ENGINEERING COMPUTATIONAL TOOLS TO STUDY AND DESIGN MOLECULAR RECOGNITION SYSTEMS

A Dissertation

by

ASUKA AUTUMN ORR

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Phanourios Tamamis
Committee Members,	Arul Jayaraman
	Efstratios N. Pistikopoulos
	Stephen Safe
Head of Department,	Arul Jayaraman

August 2021

Major Subject: Chemical Engineering

ABSTRACT

Molecular recognition comprises the noncovalent interaction of two or more binding partners and is central to many biological processes and designed agents for therapeutic or environmental applications. In this doctoral study, computational tools were engineered to address challenging in molecular recognition that are otherwise difficult to solve using conventional methods. The different computational tools comprise MD simulations, energy calculations, and structural analysis coupled with programs that strategize their execution. The tools have been developed and used to 1) elucidate and differentiate the binding of structurally and physicochemically similar ligands to proteins, 2) characterize modified RNA : protein interactions, 3) study and design affibody proteins with anti-amyloid properties, 4) examine the binding of toxic compounds onto montmorillonite clays, and 5) elucidate short-peptide self- and co-assembly. These computational tools can be considered as "*in silico* experiments" to bridge gaps between experimental observations and theory. The application of these tools have suggested potential interactions leading to biological activity and predicted stronger signaling properties of one enantiomer over the other, revealed the broader recognition of RNA binding proteins for modified RNAs, elucidated the binding and specificity of affibody proteins for amyloidogenic proteins, predicted toxic compound adsorption free energies for clays, and examined the pathways of designed peptide self- and co-assembly, which led to the discovery of novel peptide cancer drug nanocarriers with advantageous properties for bioimaging and drug delivery.

ACKNOWLEDGEMENTS

This dissertation has become a reality through the support and encouragement of some very special individuals. I am so deeply grateful for the experiences I've had over the course of my doctoral studies with the Artie McFerrin Department of Chemical Engineering at Texas A&M – both the peaks and the valleys.

I would first like to thank my advisor, Dr. Phanourios Tamamis, for his unfailing support he's given me since the day I met him, both in my professional and personal life. I'm filled with bliss to see the lab flourishing, and I am proud to have been a part of the lab's growth from its inception to today. I often think of the time you told me you had a "golden egg" project for me to work on when I told you as an undergraduate student that I would continue in your lab to pursue my PhD. Little did I know that this "golden egg" would multiply into several golden egg opportunities.

I would also like to express my gratitude to my committee members, Dr. Arul Jayaraman, Dr. Efstratios N. Pistikopoulos, and Dr. Stephen Safe. I am honored to have had you all in my PhD committee and to have worked with you as a collaborator. Your support for me has given meaning to the phrase "standing on the shoulders of giants." I hope to one day be able to be the shoulders on which others can stand as you were for me.

I owe a deep sense of gratitude to my former and present colleagues in Dr. Tamamis' lab. To Sai Vamshi Reddy Jonnalagadda, thank you for the late-night dinner runs after the long days at the lab. Thank you for being my dearest friend and ally in all my endeavors. To Joseph Jakubowski, thank you for keeping things light with the giggles

and coffee breaks. To Büşra Özgüney thank you for the discussions on theology and sending me positive vibes throughout these past two semesters. I am so excited and blessed to have you continuing my work, and I look forward to seeing all that you will undoubtedly accomplish. To all the graduate and undergraduate students of Dr. Tamamis' lab, thank you for your contributions to my studies and providing a stimulating research environment.

I would be remiss if I failed to express my heartfelt appreciation to the collaborators of Dr. Tamamis' lab. My doctoral studies would not have been possible without their support and expertise. I thank Dr. Stephen Safe, Dr. Arul Jayaraman, Dr. Lydia M. Contreras, Dr. Wolfgang Hoyer, Dr. Timothy D. Phillips, Dr. Ehud Gazit, and Dr. Lihi Adler-Abramovich as well as their lab members for our collaborative studies and their willingness to share their extensive knowledge on the studied systems. I also thank the Texas A&M High Performance Research Computing facility, which provided the advanced computing resources and consultation needed to perform my simulation studies. I would also like to thank the faculty and staff at the Artie McFerrin Department of Chemical Engineering as well as the Frederick E. Giesecke Engineering Research Building for both administrative and emotional support.

My PhD journey would not have been possible without the help of my past teachers and mentors. To all my past teachers, thank you for inspiring your students to be the best versions of themselves and reach new heights. I would particularly like to thank my high school chemistry teacher, Mr. Ken Bowen, for encouraging my interest in science and championing me in chemistry/science competitions. I know that your dedication to

students from all walks of life has positively impacted many of our lives. I am also immensely thankful to Dr. Lisa M. Pérez, who gave me my first research experience. Before joining your group, I had no experience with even basic terminal usage, let alone computational research. I gained invaluable experience from the opportunities that you granted me that eventually led me to my PhD studies. I would also like to extend my appreciation to Dr. Mustafa Akbulut and his former graduate students for providing me the resources, support, and patience they gave me for my undergraduate research in his lab. I would also like to thank Dr. Akbulut for his encouragement and endorsement to participate in the Undergraduate Summer Research Grant Program that led me to Dr. Tamamis' lab.

Finally, I am forever grateful for my family and friends for filling my life with joy and being my role models. To my parents, thank you for providing me with the strongest foundation to build my future on. Words cannot adequately describe how fortunate I feel to be your daughter. Thank you both for showing me that "the best prize that life has to offer is a chance to work hard at work worth doing." To my sister, Ashley, and Chris, thank you for always welcoming me to your home whenever I needed a place to decompress.

Thank you again to all who have been with me in my journey. I can never repay my debt of gratitude to you all, yet your kindness continues to inspire me to pay it forward and hopefully impact others at least a fraction of the way you've impacted me.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Phanourios Tamamis, Dr. Arul Jayaraman, and Dr. Efstratios Pistikopoulos of the Artie McFerrin Department of Chemical Engineering and Dr. Stephen Safe of the Department of Veterinary Physiology and Pharmacology.

A key component of this doctoral study involved collaborations with experimental labs. Experimental contributions were made by members of Drs. Safe's, Jayaraman's, Lele's, and Manson's labs at Texas A&M University, College Station, TX (Chapter 2), members of Dr. Contreras' lab at the University of Texas, Austin, TX (Chapter 3), members of Dr. Hoyer's lab at Heinrich-Heine-University, Düsseldorf, Germany (Chapter 4), members of Dr. Phillips' lab at Texas A&M University, College Station, TX (Chapter 5), and members of Dr. Gazit's and Dr. Adler-Abramovich's labs at Tel Aviv University, Tel Aviv, Israel (Chapter 6), as well as additional collaborators of the experimental collaborators listed above; the contributing experimentalists are listed as co-authors in the corresponding published studies or will be listed in future published studies. The computational work for the dissertation was completed by the student; additional computational contributions were made by members Dr. Tamamis' lab (Chapters 2-6) and Dr. Pistikopoulos' lab (Chapter 5), and contributing students are listed as co-authors in corresponding published studies or will be listed in future published studies.

Funding Sources

Graduate study was supported in part by the Graduate Diversity Fellowship from Texas A&M University. This work was also made possible in part by funding available to Dr. Tamamis, including startup funding from the Artie McFerrin Department of Chemical Engineering, the Understanding the Rules of Life: Epigenetics Program (National Science Foundation) under Grant Number 2022124 (Chapter 2), the National Institute on Aging (National Institutes of Health) under Grant Number R03AG058100 (Chapter 3), and the Superfund Hazardous Substance Research and Training Program (National Institutes of Health) under Grant Number P42 ES027704 (Chapter 4). The contents of the studies of this dissertation are solely the responsibility of the authors and do not necessarily represent the official views of the National Science Foundation or National Institutes of Health.

NOMENCLATURE

α-syn	α-Synuclein
β-wrapin	β-Wrap proteins
Aβ	Amyloid-β
AD	Alzheimer's disease
AhR	Aryl hydrocarbon receptor
AMBER	Assisted Model Building with Energy Refinement
BPA	Bisphenol A
BPS	Bisphenol S
C. crescentus	*Caulobacter crescentus*
CGenFF	CHARMM General Force Field
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CM	Ca^{2+} montmorillonite
CM-carnitine	L-carnitine-amended Ca^{2+} montmorillonite
CM-choline	Choline-amended Ca^{2+} montmorillonite
COUP-TF1/2	Chicken ovalbumin upstream promotor-transcription factor ½
cyclo-HH	cyclo-dihistidine
cyro-EM	Cryo-electron microscopy
DBP	Dibutyl phthalate
DBPs	DNA-binding proteins
DEHP	di-2-ethylhexyl phthalate

DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EMSA	Electrophoretic mobility shift assay
FRET	Förster resonance energy transfer
FF	Diphenylalanine
Fmoc-F	Fluorenylmethyloxycaronbyl-phenylalanine
Fmoc-FF	Fluorenylmethyloxycaronbyl-diphenylalanine
GenX	Hexafluoropropylene oxide
IAPP	Islet amyloid polypeptide
KH	K homology
MD	Molecular dynamics
MM–GBSA	molecular mechanics – Generalized Born surface area
MM–PBSA	molecular mechanics – Poisson Boltzmann surface area
MST	Microscale thermophoresis
NGS	Next-generation sequencing
NMR	Nuclear magnetic resonance
NOVA1	Neuro-oncological ventral antigen
PD	Parkinson’s disease
PDB	Protein Data Bank
PNPase	Polynucleotide phosphorylase
PFAS	Per- and polyfluoralkyl substances
PFBS	Perfluorobutane sulfonic acid

PFOA	perfluorooctanoate
PFOS	perfluorooctanesulfonate
QM	Quantum mechanics
RBDs	RNA binding domains
RBPs	RNA-binding proteins
REMD	Replica-exchange molecular dynamics
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
SASA	Solvent accessible surface area
SM	Na^+ montmorillonite
SPR	Surface plasmon resonance
STORM	Stochastic optical reconstruction microscopy
T2D	Type 2 diabetes
TDP-43	Transactive response DNA-binding protein 43
UV	Ultraviolet
YTH	YT521-B homology
YTHDF1	YTH N6-methyladenosine RNA binding protein 1

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGEMENTS iii

CONTRIBUTORS AND FUNDING SOURCES vi

NOMENCLATURE viii

TABLE OF CONTENTS xi

LIST OF FIGURES xiii

LIST OF TABLES xvi

CHAPTER I INTRODUCTION 1

CHAPTER II SMALL-MOLECULE LIGANDS BINDING TO PROTEINS 7

Docking-Refinement of Small-Molecule Ligand : Protein Complexes 7
Introduction 7
Methodology of the Docking-Refinement Protocol 12
Applications of the Docking-Refinement Protocol 18
Concluding Remarks on Docking-Refinement 25

CHAPTER III RNAS BINDING TO PROTEINS 26

Characterization of Modified RNA Acid : Protein Interactions 26
Introduction 26
Methodology of the Protocol for the Characterization of Modified RNA : Protein Interactions 31
Application of the Protocol for the Characterization of Modified RNA : Protein Interactions 41
Concluding Remarks for the Protocol for the Characterization of Modified RNA : Protein Interactions 49

CHAPTER IV INHIBITION OF AMYLOID SELF-ASSEMBLY 50

Study and Design of β-wrapin (Affibody Proteins) Binding and Specificity to Amyloidogenic Proteins 50

Introduction 50
Methodology for the Study and Design of β-wrapins Binding to Amyloidogenic Proteins 55
Applications for the Study and Design of β-wrapins Binding to Amyloidogenic Proteins 65
Rational Design of β-wrapins Incorporating Non-canonical Amino Acids for Bifunctional Fibril Deformation and Monomer Sequestration Properties 82
Concluding Remarks for the Study and Design of β-wrapins Binding to Amyloidogenic Proteins 87

CHAPTER V LIGANDS BINDING TO CLAYS 89

Simulation Studies of Toxic Compounds Binding to Montmorillonite Clays 89
Introduction 89
Methodology for Simulating and Analyzing the Binding of Compounds onto Montmorillonite Clays 94
Applications of Simulating and Analyzing the Binding of Compounds onto Montmorillonite Clays 105
Concluding Remarks for Simulating and Analyzing the Binding of Compounds onto Montmorillonite Clays 119

CHAPTER VI PEPTIDE-BASED SELF- AND CO-ASSEMBLY 120

Elucidating the Pathways of Peptide Self- and Co-Assembly 120
Introduction 120
Methodology for Elucidating the Pathways of Short Peptide Self- and Co-Assembly 126
Applications of Elucidating the Pathways of Peptide Self- and Co-Assembly 141
Concluding Remarks for Elucidating the Pathways of Short Peptide Self- and Co-Assembly 149

CHAPTER VII SUMMARY 150

Current and Future Perspectives 150
Docking-Refinement Protocol for Small-Molecule Ligand : Protein Complexes 150
Protocol for the Characterization of Modified RNA : Protein Interactions 152
Study and Design of β-wrapin Binding and Specificity to Amyloidogenic Proteins 154
Simulating and Analyzing the Binding of Compounds onto Montmorillonite Clays 156
Elucidating the Pathways of Short Peptide Self- and Co-Assembly 157

REFERENCES 159

LIST OF FIGURES

Page

Figure II-1. Overview of methods used for molecular docking. 9

Figure II-2. Workflow for the docking refinement protocol. 13

Figure II-3. Schematics of (A) the harmonic spherical potential and (B) the quartic spherical potential. 16

Figure II-4. Molecular graphics images of (A) TCDD and (B) 1,4-DHNA in complex with mouse AhR. 19

Figure II-5. Molecular graphics images of (A) DIM-C-Pyr4 and (B) 1,1-CH_3-DIM-Pyr4 in complex with COUPTF2. 21

Figure II-6. Molecular graphics image of the principal motions of *E. coli* Tsr in complex with (A) (*R*)-DHMA and (B) (*S*)-DHMA. 22

Figure II-7. Dynamic cross correlation maps for *E. coli* Tsr binding to (A) (*R*)-DHMA and (B) (*S*)-DHMA using Cα atom coordinates. 24

Figure III-1. Workflow for the protocol for the characterization of modified RNA : protein interactions. 32

Figure III-2. Organization of modified RNAs into trees stemming from uracil (U), cytosine (C), adenosine (A), and guanosine (G). 36

Figure III-3. Average MM–GBSA association free energies ($kcal \cdot mol^{-1}$) with respect to experimentally derived dissociation constants (nM) of RNA strands containing select modified RNAs. 43

Figure III-4. Molecular interactions of 8-oxoG with the mutant PNPases: NYM (panel A for position P8 and panel B for position P9); NYT (panel C for position P8 and panel D for position P9); NFH (panel E for position P8 and panel F for position P9); GFT (panel G for position P8 and panel H for position P9) and SYH (panel I for position P8 and panel J for position P9). 46

Figure IV-1. Overview of β-wrapin structure, sequence, and affinity to Aβ, α-syn, IAPP, and tau. 52

Figure IV-2. Workflow for the β-wrapin design strategy. 57

Figure IV-3. Molecular graphics images of the common hydrophobic interactions of AS10 in complex with (A) IAPP, (B) Aβ, and (C) α-syn. 67

Figure IV-4. Computational modeling of TP4 in complex with tau. 69

Figure IV-5. Molecular graphics image showing the improved interactions of HI18-R5-3 in complex with IAPP. 72

Figure IV-6. Molecular graphics image showing the improved interactions of HI18-R5-6 in complex with IAPP. 74

Figure IV-7. Molecular graphics image showing the improved interactions of AS10-4 in complex with Aβ, α-syn, and IAPP. 78

Figure IV-8. Molecular graphics image showing the improved interactions of AS10-2 in complex with Aβ, α-syn, and IAPP. 79

Figure IV-9. Molecular graphics images of (A) unmodified AS10 N-terminus and (B) curcumin in complex with the Aβ fibril. 86

Figure IV-10. Molecular graphics images of a (A,D) curcumin in complex with the Aβ fibril, (B,E) modified N-terminus, FDNK-(MOT5)-NKEMA, binding to the Aβ fibril, and (C,F) modified N-terminus, FDNK-(NAO2)-NKEMA, binding to the Aβ fibril (A,B,C) prior to the initiation of partial dissociation of the Aβ fibril and (D,E,F) at 100 ns. 86

Figure V-1. Schematic view of automated simulation setup for simulating the binding of toxic compounds onto montmorillonite clays. 96

Figure V-2. Molecular graphics images from representative MD simulation snapshots of (A,B) glyphosate and (C,D) paraquat binding to SM at (A,C) pH 2 and (B,D) pH 7. 107

Figure V-3. Molecular graphics of the simulation snapshots containing representative and prominent binding modes for (A) PFOA and (B) PFOS binding to CM, and (C) PFOA and (D) PFOS binding to CM-carnitine. 110

Figure V-4. Correlation of the experimentally observed adsorption free energies (x-axis) and the computationally predicted adsorption free energies (y-axis) of compounds binding to CM (α = 1.01 and β = 0.48). 115

Figure V-5. Molecular graphics image of chemicals encompassing characteristics linked to CM adsorption. 118

Figure VI-1. Schematic of the overall computational methodology to elucidate the pathways of short peptide self- and co-assembly. 126

Figure VI-2. Overview of the formation of clusters by cyclo-HH in the absence and presence of Zn^{2+}. 143

Figure VI-3. Overview of the formation of clusters by cyclo-HH in the presence of $Zn(NO_3)_2$ and Epirubicin. 145

Figure VI-4. Structural properties of the clusters formed by Fmoc-3,4F-Phe and Fmoc-3,5F-Phe peptides within MD simulations. 148

LIST OF TABLES

Page

Table IV-1. Summary of the most promising designed β-wrapin variants for IAPP.....70

Table IV-2. Summary of the most promising designed β-wrapin variants for multi-targeted Aβ, α-syn, and IAPP binding.....76

Table IV-3. Summary of the most promising designs β-wrapin variants for multi-targeted Aβ and tau binding.....81

CHAPTER I

INTRODUCTION

Molecular recognition refers to the noncovalent interaction of two or more molecular binding partners.[1] Such binding partners include, among others, different proteins, proteins and nucleic acids, proteins and small-molecules, different small-molecules, as well as inorganic/organic surfaces and small-molecules. Molecular recognition is central to many biological processes.[2-18] A subset of these abundant processes include: The intricate coordination of RNAs and RNA-binding proteins (RBPs) involved in RNA-metabolism, translation, and gene regulation at both the transcriptional and post-transcriptional levels;[2-8] Small-molecule ligands binding to proteins to act as coenzymes or substrates in many basic enzymatic reactions as well as extra- and intra-cellular signals to help construct regulation networks;[9-16] Proteins self-assembling into crystals, gels, filaments, among other varieties of aggregates to form structures including virus capsids[17] and amyloid fibrils.[18] Additionally, molecular recognition is a key component of several agents or materials for therapeutic[19-22] or environmental[23-25] applications.

The use of computers to study and predict the properties and structures involved in molecular recognition has become increasingly significant in the past decades.[26-28] The advent and continual improvement of high-performance computing has enabled *in silico* experimentation bridging gaps between observations from laboratory experiments and theory. Several computational methods have been developed to study, understand, and

enhance molecular recognition systems.[26, 27, 29-32] One such computational tool, molecular dynamics (MD) simulations, has been widely used as a "computational microscope"[33] to understand the structural properties, dynamic evolution, and the function of molecular recognition systems.[28, 33, 34] Importantly, numerous studies[35-43] have demonstrated that MD simulations can complement and resolve ambiguity in experimental data by extending static structural data into time-evolved snapshots of atomic-coordinates or "movies" of how molecules behave over time (reviewed in refs. [28, 33, 34]). Furthermore, insights gained from MD simulations have also provided impetus for experimental studies (reviewed in ref. [34]).

Analysis of MD simulation results is important to understand the properties of the simulated systems.[34] One simulation could track the positions and velocities for several hundreds of thousands of atoms over millions of time steps. Thus, parsing through these data to focus on the most relevant and biologically important aspects of simulation data can be formidable task, and in many cases, the most informative quantitative data are difficult to specify in advance. Identifying the maximally informative data in simulations often requires a methodical balance between visual analysis and quantitative analysis. However, once the maximally informative simulation data-based metrics are identified, simulations-based methods can provide substantially more accurate energetic and structural estimates that other computational approaches.[38, 44, 45] Free energy perturbation and other "alchemical" energy calculation methods in which a ligand or residue is gradually "transformed" into another through a series of simulations generally offer the most accurate estimates of binding free energies.[34, 46] These methods are computationally

expensive yet sufficiently reliable.[47, 48] The molecular mechanics – generalized Born surface area (MM–GBSA)[49] and molecular mechanics – Poisson Boltzmann surface area (MM–PBSA)[49] methods, which use continuum solvent models rather than an explicit representation of the solvent, offer substantially less computationally intensive, albeit less accurate, binding free energy estimates.[50] These energy calculation methods can be combined with structural analysis to describe the simulation systems including key interactions acting as switches promoting signaling, energetic favorability of pathways leading to protein aggregation or folding, and ligands recognized by a protein receptor.

The overarching theme of this doctoral study is the development of different computational tools that can serve as bridges to solve challenging problems that are otherwise difficult to solve using existing experimental or computational approaches alone. The individual areas of the doctoral study are organized into five main chapters:

1. Chapter 2 is dedicated to the investigation of how structurally and physicochemically similar small-molecule ligands bind to proteins. Determining and differentiating the structure of structurally and physicochemically similar small-molecule ligands bound within a protein structure is challenging experimentally,[51, 52] and conventional docking methods cannot perform successful prediction for similar ligands.[53] To understand how subtle structural and physicochemical differences in ligands can lead to differences in their functional properties upon binding to a protein receptor, a docking-refinement tool that can delineate the binding of similar ligands was developed.[54-59]

2. Chapter 3 is dedicated to the investigation of which modified RNAs bind to proteins how they are recognized. Over 150 chemically modified RNAs have been identified,[60] yet their effect on the function of RNA : protein complexes has been largely unexplored[61] due to major challenges in mapping the epitranscriptome.[62-64] To uncover the repertoire of modified RNAs that can be recognized by a target protein, a suite of programs that screen and detect modified RNAs prone to interact favorably with a target protein were developed.[65-67]
3. Chapter 4 is dedicated to the study and design of affibody proteins, termed β-wrapins, binding to amyloidogenic proteins. Few, subtle changes at the sequence level of β-wrapins greatly change their affinity to amyloidogenic proteins involved in Alzheimer's disease, Parkinson's disease, and type 2 diabetes.[68-73] Previous attempts to enhance β-wrapins' single-targeted and multi-targeted binding to amyloidogenic proteins have greatly relied on phage display library experiments and researcher intuition.[68-73] To study and design β-wrapin variants for enhanced single-targeted and multi-targeted binding to amyloidogenic proteins, MD simulations and free energy calculations were used to uncover the binding and specificity of β-wrapins,[74, 75] and an optimization-based protein design program accounting for both single-targeted and multi-targeted binding was developed. Rationally designed β-wrapins with N-termini modified with non-canonical amino acids were also designed to mimic the mechanism of action of curcumin binding to Aβ fibrils and inhibiting amyloid elongation.

4. Chapter 5 is dedicated to the investigation of compound adsorption to surfaces. Montmorillonite clays are a promising adsorbent for several, diverse toxic compounds to mitigate their toxicity.[24, 25] To determine the binding mechanisms of compounds to clay surfaces, the generation of MD simulation input files and parameters was automated to systematically examine the binding properties of compounds to adsorbent surfaces.[76-78] To screen compounds for which montmorillonite clays can be effective adsorbents, a combination of minimalistic simulations, simple interaction energy calculations, and a minimalistic model was developed to predict the adsorption free energies of diverse compounds for montmorillonite clay.[79]
5. Chapter 6 is dedicated to the investigation of short peptide self- and co-assembly. The initial stages of peptide self-assembly is difficult to probe experimentally due to the complex nature of self-assembly and limitations due to sensitivity and time resolution.[80] To examine the initial stages of short peptide self- and co-assembly and the effect of different solvents, ions, or modifications to the peptides, a suite of structural and energetic analysis programs tracking and characterizing the formation of clusters within MD simulations was developed.[81-84]

The term molecular recognition is used in this doctoral study to refer to the interactions between the different molecular bodies of each individual chapter. Molecular recognition processes are involved in the binding and signaling of small-molecule ligands in complex with proteins[85-87] (Chapter 2), the preferential binding of RNAs in complex with proteins[88-92] (Chapter 3), the sequestration of proteins by affibody proteins[68-70, 72, 73,

[93-95] (Chapter 4), the tight binding of specific types of toxic compounds onto clay surfaces[96-100] (Chapter 5), and the gradual formation of interactions leading to the organization and stabilization of self- and co-assembled short-peptides, molecules, and/or ions into specific geometries and nanostructures[101-106] (Chapter 6). The common basis and governing principle associated with providing insights into and solving these problems comprise the use of MD simulations and energy calculations coupled with FORTRAN programs that strategize the execution of MD simulations and energy calculations as well as perform the structural and energetic analysis of the systems under investigation, in tandem with existing programs and tools in the literature.

CHAPTER II

SMALL-MOLECULE LIGANDS BINDING TO PROTEINS

Docking-Refinement of Small-Molecule Ligand : Protein Complexes

Introduction

Delineating the structures of small-molecule ligand : protein complexes is a key problem due to its importance in the molecular understanding of biological phenomena and provide insights enabling drug discovery and design.[107] The development of experimental techniques including high-throughput protein purification, crystallography, and nuclear magnetic resonance (NMR) spectroscopy contributed valuable insights into the structures of a vast number of small-molecule ligand : protein complexes. Despite advances in experimental structure determination techniques, the process is often time consuming and expensive with relatively large quantities and high concentrations of the molecules required.[108] Additionally, experimentally determining and differentiating the structure of structurally and physicochemically small-molecules bound within a protein structure can be challenging.[51, 52] A higher thermal motion or conformational disorder of the small-molecules in comparison with the surrounding protein can lead to less well-defined electron density.[51] Further, standard crystallographic assessment tools face challenges in determining the protonation state, tautomeric state, stereochemistry, orientation, and/or conformation of the small-molecule as multiple or all the different states of the small-molecule may fit within the electron density.[109]

Molecular docking is a widely used computational approach capable of both identifying probable conformations of a ligand binding to a target protein and estimating the strength of interaction between a ligand and a target protein.[110-114] Computational molecular docking has been implemented in several studies to predict how small-molecules induce/prevent signaling upon binding to a target protein[110, 113, 115, 116] and in structure-based drug discovery.[117-119] Computational docking approaches rely on the combination of a search algorithm to generate possible binding conformations and a scoring function to evaluate the generated conformations.[56] Advances in computational power has allowed the evolution of docking algorithms to include a wide array of search algorithms and scoring functions.[117] Despite the successes of current docking methods in providing fundamental knowledge[110, 113, 115, 116] as well as discovering novel drugs,[117-119] a comprehensive study based on an extensive dataset of 2002 protein : ligand complexes extracted from the PDBbind database (version 2014)[120] showed that, while the correct ligand binding poses could be generated in most cases by the evaluated docking programs, the ranks of binding affinities for the entire dataset could not be well predicted by most docking programs.[57] Additionally, the evaluation of the docking programs showed that each individual docking program had different scoring powers for different protein families, and therefore the use of different docking programs for different protein families was suggested.[57]

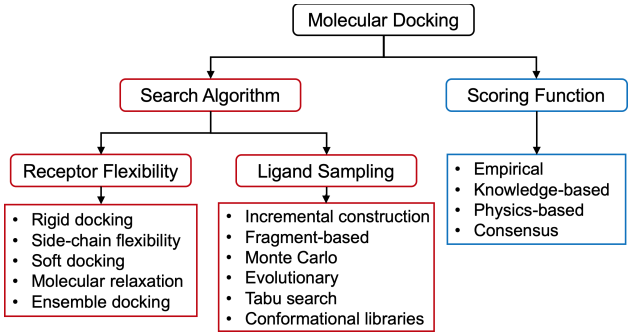

Figure II-1. Overview of methods used for molecular docking.

The procedure for computational molecular docking may be categorized into two principal stages: (1) identification of potential ligand-binding pockets of the receptor, and (2) identification of the most energetically favorable binding conformation in the identified binding pockets[54] (Figure II-1). Experimental information that can reveal a ligand binding site of a target protein, such as site-directed mutagenesis studies, can be used to circumvent the first stage of molecular docking. In such a case, the search space can be confined to a ligand binding site of a target protein, and a docking program can be introduced to determine the orientation and position of a ligand within the receptor's binding pocket. The reduction of search spaces based on information regarding the binding site location can be advantageous in increasing the docking efficiency.[57]

Improvements in computational power and more efficient docking practices (e.g. searching a specific binding site rather than the entire receptor) has allowed docking methods to graduate from the "lock-and-key" assumption, treating both the ligand and receptor can as rigid bodies,[58] to the "induced-fit" theory suggesting that the ligand and receptor should be treated as flexible during docking.[59,60] Docking algorithms that allow for flexibility in the ligand and/or receptor not only predict the binding mode of a ligand

with higher accuracy than rigid docking algorithms, but also the binding affinity of a ligand relative to other compounds.[61]

Docking refinement approaches combine the benefits of rigid and flexible docking by first docking the ligand to the receptor in a high-throughput fashion and later refining the identified binding modes through molecular dynamics or local resampling strategies to provide more accurate relative binding affinity estimations and binding mode predictions.[119, 121-126] Docking refinement has proven to be an effective approach.[119, 121-126] In such methods, MD simualtions can be used used to refine the docked structure,[119, 122-124] and local resampling strategies can be used to maintain the same positioning and orientation of the small-molecule while resampling portions of the docked molecule and/or protein side chains.[125-127]

In this chapter, the methodology and applications of an in-house docking-refinement protocol developed as part of this doctoral study is described. In contrast to other docking-refinement approaches in which high-throughput docking programs are used to determine the general orientation and position of the small-molecule,[119, 121-126] MD simulations are used to explore different binding modes of a ligand within an unconstrained binding pocket, providing significantly greater flexibility to the complex structure in the exploration of binding modes. The docking-refinement protocol consists of a multifaceted pipeline, using docking simulations and short MD simulations to identify the most favorable conformations for further analysis. The combined use of several computational strategies was inspired by previous studies combining protein-protein docking, MD simulations, and free energy calculations to investigate the molecular

recognition of chemokine receptors by chemokines or protein fragments of HIV-1,[128-130] as well as C5aR by PMX53.[131] The use of such a docking-refinement protocol is particularly important in the systems of this doctoral study in which small-molecule ligands with subtle structural and physicochemical differences induce different signaling properties.

Methodology of the Docking-Refinement Protocol

The docking-refinement protocol can be outlined as follows: (1) the ligand is initially positioned in the target receptor's binding pocket based on available experimental data; (2) multiple short MD docking simulations of the ligand within the binding pocket are performed, in which the ligand is constrained in the pocket by harmonic or quartic spherical potentials and is forced to rotate and sample different binding modes in the binding pocket with binding pocket residues unconstrained; (3) the sampled binding modes generated by the short docking simulations are assessed through interaction-energy calculations to select the most probable binding modes; (4) the selected most probable binding modes are investigated through explicit-solvent MD simulations and physical-chemistry-based free energy calculations to identify the most favorable binding mode of the small-molecule ligand : protein receptor complex (Figure II-2). In this section, the stages of the docking-refinement protocol are described. The protocol is also detailed in ref. [54].

Stage 1: Initial Positioning of the Ligand in the Target Protein's Binding Pocket

The protocol uses a preliminarily docked structure of the ligand within the known binding site of the target protein as an initial structure for subsequent docking simulations, described in the following section (Stage 2). The ligand under investigation is preliminarily docked into its known or suspected binding site based on experimental data. For example, if a bound structure of the receptor is available, and if the ligand under investigation is known to bind to or interact with residues of the same binding pocket, then

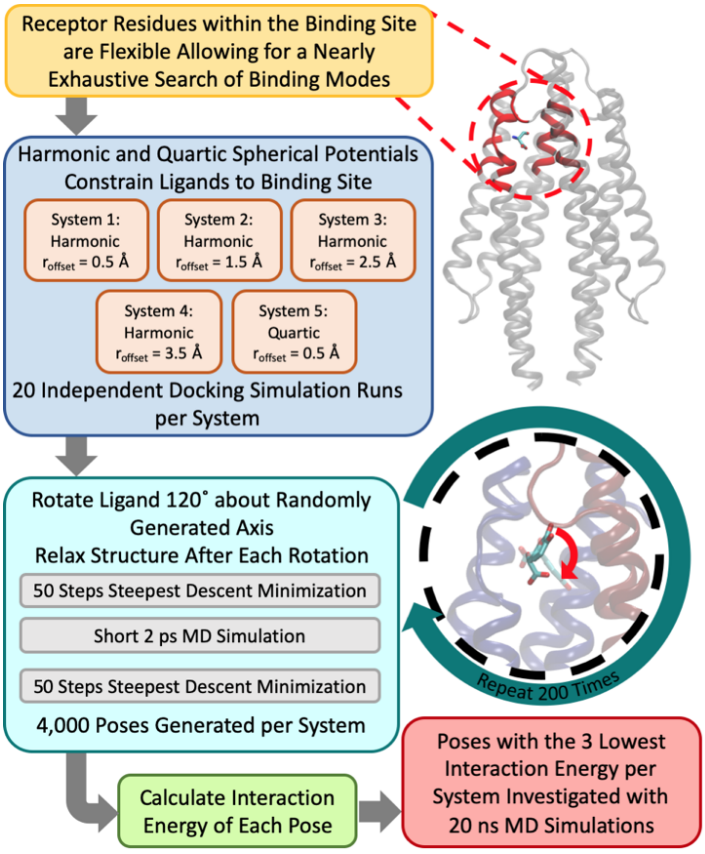

Figure II-2. Workflow for the docking refinement protocol. This figure is adapted from ref. [55]*.

the initial placement of the ligand under investigation can be guided by the location of the resolved ligand in complex with the receptor as in refs. [54, 55, 132]. Otherwise, if a bound structure is unavailable, then high throughput docking programs with the search space

* Reprinted with permission from "Molecular Mechanism for Attractant Signaling to DHMA by E. coli Tsr" by Asuka A. Orr, Jingyun Yang, Nitesh Sule, Ravi Chawla, Kenneth G. Hull, Mingzhao Zhu, Daniel Romo, Pushkar P. Lele, Arul Jayaraman, Michael D. Manson, and Phanourios Tamamis, 2020. *Biophysical Journal*, 118 (2), 492-504, Copyright 2020 by Elsevier.

optimally guided by experimental data and mutagenesis studies determining receptor residues involved in ligand binding can be used as in refs. [56-59]. Although the initial placement is not explicitly considered in the analysis and does not directly contribute to the conformations generated in the later steps, the initial placement is important as it determines the initial position of the center of mass of the ligand, and in this way, it influences the generated binding poses.

Stage 2: Short Docking Simulations Nearly Exhaustively Searches Binding Modes

After the small-molecule ligand under investigation has been initially placed within the binding site of the protein receptor, short docking simulations are introduced to search possible binding modes nearly exhaustively. In the docking simulations, 20 independent runs are performed in which 200 steps of 2 ps MD simulations are conducted. For each of the 200 steps, the ligand is rotated about a randomly generated axis followed by a 2 ps MD simulation and energetic minimization (Figure II-2, cyan block); the final conformation after the short simulation and energetic minimization is saved for subsequent evaluation. Thus, for each docking simulation consisting of 20 independent runs, 4000 binding modes of the ligand under investigation in complex with the protein receptor are saved.

Separate sets docking simulations, or docking systems are introduced to explore the possible binding modes of the ligand (Figure II-2, orange blocks within blue block). The docking systems are distinct from each other based on how the ligand is constrained in the binding site. A subset of docking systems constrains the ligand through harmonic

spherical potentials (Figure II-1A), while the remaining docking systems constrain the ligand though quartic spherical potentials (Figure II-1B). The quartic spherical potential introduces an energetic well away from the center of the spherical potential; this encourages the ligand to explore binding modes from its initial positioning and reduce the bias of the initial placement (Figure II-1B). Additionally, for each docking system, an r_{offset} is defined such that the energy potential is zero if the difference between the initial and new (after the rotation around a random axis) center of mass of the ligand is less than r_{offset} (Figure II-2, orange blocks within blue block). Within the short docking simulations, both the side chain and backbone atoms of binding site receptor residues, which are defined as residues with Cα atoms within a user specified distance cutoff of the initially placed ligand (Figure II-2, yellow block), are unconstrained and flexible. The absence of constraints on the binding site residues aims to facilitate the ability of the residues to adapt to the different binding modes of the ligand generated through random rotations. Additionally, the Generalized Born with a Simple Switching (GBSW) implicit solvent model[133] is used in the docking simulations to further facilitate the docking procedure; the absence of explicitly modeled solvent molecules increase the flexibility of the binding site and decrease the computational resources needed for the simulations to complete.

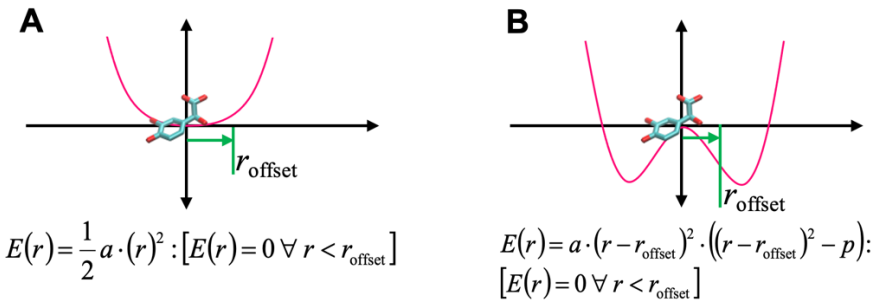

Figure II-3. Schematics of (A) the harmonic spherical potential and (B) the quartic spherical potential.

Stage 3: Initial Screening of Binding Modes through Interaction Energy

For each docking system, the interaction energy of the 4000 generated binding poses of the small-molecule ligand in complex with the protein receptor, generated in Stage 2, is calculated (Figure II-2, green block). The three binding modes with the lowest interaction energies (corresponding to the sum of van der Waals and electrostatic interaction energies between the ligand and the protein) per system are saved for subsequent evaluation in Stage 4. This step acts as a quick initial screening in which the binding poses with the lowest interaction energies are selected for further investigation.

Stage 4: Refinement and Final Evaluation of Selected Binding Modes

All-atom explicit solvent MD simulations are performed using the binding modes selected in Stage 3 as starting conformations (Figure II-2, red block). In this stage, multi-ns simulations are introduced to refine the ligand : protein conformations, improve intermolecular interactions, and ultimately to assess the most energetically favorable binding mode. Upon completion of the multi-ns simulations, the energetic favorability of

the binding modes is assessed using MM–GBSA[49] association free energy calculations. The combined use of MD simulations with physics-based calculation of free energy have proved beneficial for detecting the lowest binding free energy mode of several ligand (protein, peptide, or small-molecule) : protein complexes. Thus, the binding modes are compared based on their average MM–GBSA[49] association free energies, averaged over the number of simulation snapshots. Under the governing principle that the most energetically favorable binding mode is the naturally occurring binding conformation, the simulation trajectory encompassing the most energetically favorable binding mode is selected. A highly detailed trajectory analysis of the selected lowest average MM–GBSA[49] association free energy simulation(s) can be performed to provide valuable insights into ligand : protein complex as described in the following section.

Applications of the Docking-Refinement Protocol

The docking-refinement protocol has been used in several studies to detect the most energetically favorable binding modes of small-molecule ligands binding to a diverse set of protein receptors. The protocol has been used to suggest key interactions for compounds binding to the human or mouse AhR,[56-58] compounds binding to human COUP-TF1 and COUP-TF2,[59] as well as compounds binding to *Escherichia coli* (*E. coli*) Tsr.[54, 55] Experiments for these studies were performed by members of Drs. Safe's, Jayaraman's, Lele's, and Manson's labs at Texas A&M University, College Station. Contributors to these studies are listed in the authors lists of refs. [54-59]. Additional details of these study are provided in refs. [54-59].

Aryl Hydrocarbon Receptor Interactions with Agonists and Antagonists

The aryl hydrocarbon receptor (AhR) is a ligand-activated transcription factor that has been shown to regulate the immune response and is a key component in protecting the gut against inflammation and cancer (reviewed in ref. [68]). In collaboration with Dr. Safe's lab, the binding of microbial metabolites[56] as well as flavonoids and isoflavones[57, 58] to AhR was investigated using the docking refinement protocol. Figure II-4 shows the key interactions of TCDD and 1,4-DHNA binding to mouse AhR. Similarities and differences in the strength of interaction of agonist versus antagonist ligands to AhR residues were used to suggest interactions that may act as switches inducing agonist activity. These studies suggested that polar interactions with Ser336/330 (residue numbers correspond to human AhR/mouse AhR numbering) and Gln383/377 as well as hydrophobic interactions

with at least a portion of residues Ile325/319, Cys333/327, Met348/342, Phe351/345 could be important for an agonist ligand's functional properties.[56-58] It is worth noting that studies of other diverse ligands within the AhR binding domain also suggested the specific combinations of residues could play a critical role in the ligands' binding and functionality.[134-139]

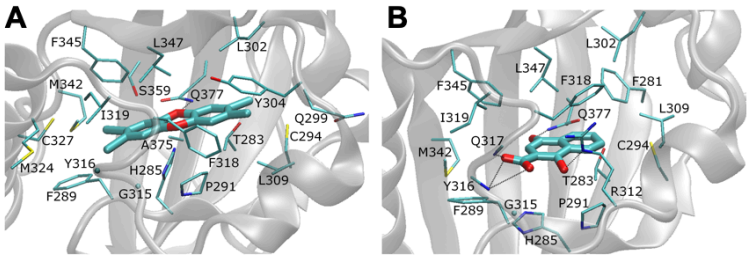

Figure II-4. Molecular graphics images of (A) TCDD and (B) 1,4-DHNA in complex with mouse AhR. This figure is adapted from ref. [56*].

* Reprinted with permission from "Editor's Highlight: Microbial-Derived 1,4-Dihydroxy-2-naphthoic Acid and Related Compounds as Aryl Hydrocarbon Receptor Agonists/Antagonists: Structure–Activity Relationships and Receptor Modeling" by Yating Cheng, Un-Ho Jin, Laurie A Davidson, Robert S Chapkin, Arul Jayaraman, Phanourios Tamamis, Asuka Orr, Clint Allred, Michael S Denison, Anatoly Soshilov, Evelyn Weaver, Stephen Safe, 2016. *Toxicological Science*, 155 (2), 458-473, Copyright 2016 by Oxford University Press.

Chicken Ovalbumin Upstream Promoter-Transcription Factor I Activation by Ligand Binding

Chicken ovalbumin upstream promoter-transcription factors (COUP-TFs) are orphan members of nuclear receptor superfamily that activate or repress gene transcription by directly binding DNA sequence.[69] COUP-TF1 and COUP-TF2 proteins are 95% homologous and evolutionarily conserved in the DNA binding domain, mainly differing at the N-terminus (reviewed in ref. [69]). In collaboration with Dr. Safe's lab, modeling studies were performed using the docking refinement protocol to elucidate the interactions of DIM-C-Pyr4 and 1,1-CH_3-DIM-Pyr4 within the ligand binding domain of COUP-TF1/2.[59] Despite their similar chemical structures, the modeling studies suggested that the ligands adopt different binding modes in complex with COUP-TF1/2 due to the fact that the methyl groups of 1,1-CH3-DIM-Pyr4 impede its ability to form hydrogen bonds with the backbone carboxyl group of V384/377, which occurs between DIM-C-Pyr4 and COUP-TF1/2. Instead, in 1,1-CH_3-DIM-Pyr4, the methyl groups are attracted to COUP-TF1/2 residue motif 219/212–227/220. Due to the somewhat different binding modes of the two compounds, only 1,1-CH_3-DIM-Pyr4 (inactive in transactivation) interacted strongly with COUP-TF1/2 residues I219/212, C220/213, A223/216, L227/220, F295/288, S383/376, S381/374, S382/375, F383, and V394/387 (Figure II-5B); while only DIM-C-Pyr4 interacted strongly with COUP-TF1/2 residues S257/250, V261/254, F302/295, V380/373, V384/377, I385/378, and F389/382 (Figure II-5A). These differences could influence their different activities as activators of COUP-TF1.[59]

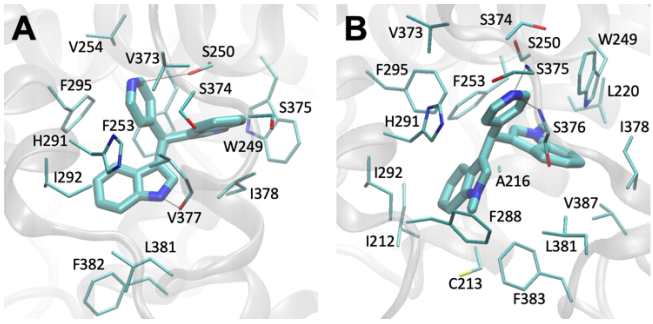

Figure II-5. Molecular graphics images of (A) DIM-C-Pyr4 and (B) 1,1-CH_3-DIM-Pyr4 in complex with COUPTF2. This figure is adapted from ref. [59*].

Mechanism of DHMA-Induced Attractant Signaling by *E. coli* Tsr

The attractant chemotaxis response of *Escherichia coli* (*E. coli*) to norepinephrine involves the conversion of norepinephrine into 3,4-dihydroxymandelic acid (DHMA) and the interaction of DHMA to the *E. coli* Tsr chemoreceptor.[70] Initial studies used a racemic mixture of the (*R*) and (*S*) enantiomers, leaving open the question of which chiral form is active. To investigate the binding properties of (*R*)-DHMA and (*S*)-DHMA to *E. coli* Tsr, the docking-refinement protocol in combination with multiple multi-ns MD simulation, free energy calculations, and structural analysis were used.[54, 55]

The docking refinement protocol suggested that (*R*)- and (*S*)-DHMA bound to Tsr with slightly different orientations within the serine-binding site.[54, 55] Additional simulations and energetic/structural analysis of the docked structures predicted that (*R*)-DHMA should promote a stronger attractant response than (*S*)-DHMA due to a

* Reprinted with permission from "Activation of COUP-TFI by a Novel Diindolylmethane Derivative" by Kyungsil Yoon, Chien-Cheng Chen, Asuka A. Orr, Patricia N. Barreto, Phanourios Tamamis, and Stephen Safe, 2019. *Cells*, 8 (3), 220, Copyright 2019 by MDPI.

consistently greater-magnitude piston-like pushdown of the Tsr α-helix 4 toward the membrane upon binding of (*R*)-DHMA than upon binding of (*S*)-DHMA (Figure II-6).[55] Tsr α-helix 4 is directly contiguous with the transmembrane helix, which connects the periplasmic Tsr domain to the cytoplasmic HAMP domain,[140, 141] and X-ray studies noted that a conformational change in this α-helix is induced by the binding of *L*-serine,[142] a known Tsr chemoattractant.[143] Additionally, the chemotaxis signaling of Tar by aspartate was suggested to be induced by similar piston-like motions.[115]

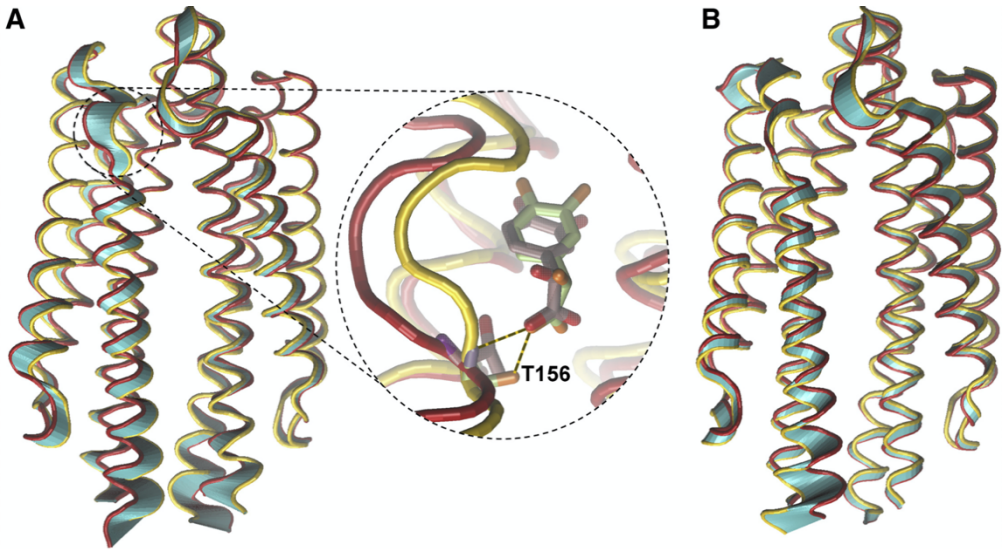

Figure II-6. Molecular graphics image of the principal motions of *E. coli* Tsr in complex with (A) (*R*)-DHMA and (B) (*S*)-DHMA. The motion between the relaxed (red) and pushed-down (gold) states of α-helix 4 is represented by the cyan tube representation. This figure is adapted from ref. [55*].

---

* Reprinted with permission from "Molecular Mechanism for Attractant Signaling to DHMA by E. coli Tsr" by Asuka A. Orr, Jingyun Yang, Nitesh Sule, Ravi Chawla, Kenneth G. Hull, Mingzhao Zhu, Daniel Romo, Pushkar P. Lele, Arul Jayaraman, Michael D. Manson, and Phanourios Tamamis, 2020. *Biophysical Journal*, 118 (2), 492-504, Copyright 2020 by Elsevier.

Dynamic cross correlation calculations showed that the conformational change in α-helix 4 was correlated with motions within the binding pocket (Figure II-7), suggesting that differences within the binding pocket induced the allosteric differences between the (*R*)- and (*S*)-DHMA bound Tsr.[55] Further investigation showed that the interaction of DHMA with Tsr residue Thr156, which had been shown by genetic studies to be critical for the attractant response to *L*-serine and DHMA[70] (Figure II-6A), was the primary cause for the piston-like motions.[55] These findings provided impetus to experimentally separate the two chiral species and test their chemoattractant strengths. Both tethered cell and motility migration coefficient assays performed by experimental collaborators validated the computational prediction that (*R*)-DHMA is a stronger attractant than (*S*)-DHMA.[55] This study demonstrated that the docking-refinement protocol and simulation studies combined with experiments can be used to investigate situations in which subtle differences between ligands may lead to diverse chemotactic responses.[55]

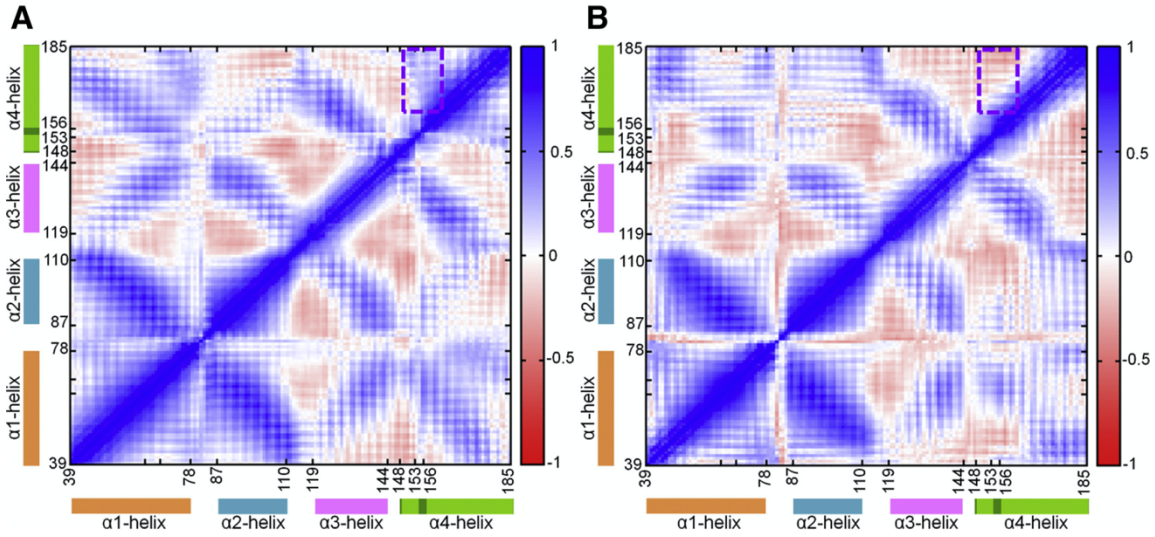

Figure II-7. Dynamic cross correlation maps for *E. coli* Tsr binding to (A) (*R*)-DHMA and (B) (*S*)-DHMA using Cα atom coordinates. The color range to the right of each map indicates the degree of correlation (or anticorrelation), with blue indicating fully correlated motions and red indicating fully anticorrelated motions. The orange, cyan, pink, and green blocks highlight residue motifs belonging to α-helix 1, α-helix 2, α-helix 3, and α-helix 4 of Tsr, respectively. The region encapsulated by green dotted lines denotes the key difference in correlated motions between Tsr bound to (*R*)-DHMA versus (*S*)-DHMA. This figure is adapted from ref. [55*].

Concluding Remarks on Docking-Refinement

In this chapter, an in-house docking-refinement protocol was developed and applied to elucidate the binding of small-molecule ligands with subtle structural and physicochemical differences to various protein receptors.[54-59] Docking MD simulations are used to explore different binding modes of a ligand within an unconstrained binding pocket, providing significantly greater flexibility to the complex structure in the exploration of binding modes. During the docking simulations, the independent use of harmonic and spherical potentials constraining the ligands within the protein's binding pocket enables the exploration of binding modes both away from (quartic) and in proximity to (harmonic) their initial position. Additionally, the use of both short MD simulations in the docking simulations and the explicit-solvent MD simulations introduce flexibility to the protein and allow for the refinement of the initial structures of the candidate binding modes before selecting the most energetically favorable ones for further investigation.[54-59] The docking-refinement protocol has enabled the study and differentiation of physicochemically and structurally similar ligands binding to several proteins.[54-59] The ability to differentiate the binding of ligands with different functional properties has allowed for the delineation of potential key interactions for the ligand's functional properties and provide a deeper molecular understanding of the protein's functional mechanisms.[54-59]

CHAPTER III

RNAS BINDING TO PROTEINS

Characterization of Modified RNA Acid : Protein Interactions

Introduction

Nucleic acids participate in nearly every cellular process and serve vital functional roles in the regulation of gene expression.[3, 4, 144, 145] Many of these vital functions rely on the intricate coordination of RNAs with RNA-binding proteins (RBPs).[2-6, 8, 146] Due to the significance of RNA : protein interactions, [2-6, 8, 146] several experimental methods have been developed to study RNA–protein interactions including electrophoretic mobility shift assay (EMSA),[147] florescent anisotropy/polarization,[148] Förster resonance energy transfer (FRET),[149] surface plasmon resonance (SPR),[150] microscale thermophoresis (MST),[151, 152] and stochastic optical reconstruction microscopy (STORM).[153] These powerful methods have allowed experimentalists to assess the specificity and affinity of RNA : protein interactions, but such methods are limited to the study of single or few molecular interactions at a time.[6] Later, more recent developments have also given rise to large-scale quantitative methods based on next-generation sequencing (NGS)[154, 155] and mass spectrometry,[156] contributing to the genome-wide identification of RBPs, RNA targets, and cofactors.[4]

A more comprehensive understanding of the fundamental mechanisms of nucleic acid-protein interactions can be gained through the determination of the three dimensional (3D) atomic structures of RNA : protein complexes.[157] Recent advances in X-ray

crystallography, solution nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryo-EM) techniques have enabled the determination of 3763 3D RNA : protein structures deposited in the Protein Data Bank (PDB)[158] (as of 5/20/2021); yet, despite the power of such methods, determining complex structures present challenges in specific cases due to difficulties in crystalizing the structures, regions of disordered electron density, poor crystallization efficacy, modeling constraints, or disordered solvent molecules within the structure.[159] Thus, computational structure prediction methods have been developed to significantly facilitate our knowledge of biomolecular interactions. During the last decades, the understanding of RNA : protein interactions has been enriched by the use of computational methods to predict RNA binding sites using homolog structures or template-based methods,[160-164] by docking RNAs to proteins,[165-169] and by simulating RNA : protein interactions.[39, 43, 66, 170-176] To this end, all-atom molecular mechanics topologies and parameters compatible with existing force fields describing nucleic acids have been developed,[42, 177-179] providing the means for researchers to investigate interactions between nucleic acids and proteins at an atomistic level.[39, 43, 66, 170-176] Apart from studying and predicting the conformation of RNA structures[180-183] and RNA : protein complexes,[39, 43, 66, 170-176] computational tools have also been useful in predicting potential protein binders[184-186] and in designing proteins to bind specific RNAs.[176, 187-189]

A particular subject of interest is the study of modified RNAs and their interactions to proteins. The presence of chemical modifications to canonical RNAs can influence the functions of RNA : protein interactions.[190] Over 150 highly diverse modified RNAs have

been identified[60] with possible functional roles in cellular metabolism.[190] Despite the fact that technological advances investigating RNA : protein interactions have recently ignited interest in the field of modified RNAs, little is known on how modified RNAs affect the versatility of RNA function in varying environmental conditions[191] and on their effect on the function of RNA : protein complexes.[61] Even with the advances in sequencing technologies and experimental methods investigating RNAs binding to proteins, several major challenges still exist for mapping the epitranscriptome (reviewed in refs. [62-64]). Most of the recently developed methods to detect and profile modified RNAs are low-throughput, based on the coupling of next-generation sequencing (NGS) technologies[62, 64] to biochemical RNA immunoprecipitation approaches or to chemical conversion reactions of specific modifications and result in ribonucleoside misincorporation patterns due to reverse transcriptase arrest; in these cases the value of any RNA immunoprecipitation data crucially depends on the quality of the antibody and sequencing accuracy.[62] However, the major challenge of NGS-based methodologies is that they are "blind" to the coexistence of several neighboring modifications (identical or different) on the same RNA molecule.[62] They also lack the resolution to detect the stoichiometry of modified RNAs.[62] In addition, the reliability of the large amount of data produced by NGS cannot be verified independently or easily by alternative methods.[62]

To understand the repertoire of the chemical modifications that a protein potentially recognizes, an independent screening experiment for each experimentally confirmed RNA chemical modification, resulting in ~150 screening experiments that would be required, involving the chemical synthesis of the modified RNAs and a

considerable number of biochemical binding assays. It should be noted that, even commercially, many of these chemical modifications may be difficult to synthesize or may not yet have procedures developed for their formation. For this reason, interdisciplinary approaches combining experimental and computational approaches, which can provide a less costly means to screen modifications prior to experimentation, constitute a promising avenue for structure-based research, design, and prediction of interactions between chemically modified RNAs with proteins. While the importance of chemical modification to molecular recognition has been captured by the number of computational approaches that exist for the study and design of protein-protein interactions with unnatural/post-transcriptionally modified amino acids (e.g. refs.[178, 192]), prior to these doctoral studies, to our knowledge, there were no reported computational structure-based protocols for the study of chemically modified RNA : protein interactions with a wide spectrum of possible modified RNAs.

Until recently, the only available molecular mechanics parameters to study such interactions comprised of the 107 modified ribonucleotides available through AMBER-compatible force field parameters.[178] These parametrization efforts focused on the atomic partial charges, and relied on analogy with similar functional groups in the AMBER force field for all other parameters.[178] Recently, MacKerell's and Nilsson's group have determined molecular mechanics force field parameters compatible with the CHARMM36 all-atom additive force field for 112 modified ribonucleotides using the CHARMM force field parametrization protocol;[177] their approach emphasizes fine-tuning of partial atomic charges and torsion angle parameters. This approach uses extensive quantum mechanics

calculations and molecular dynamics (MD) simulations of ribonucleotides in aqueous solutions for refinement against experimental data.[177] Also, the capabilities of the CHARMM General Force Field (CGenFF),[193-196] a program that performs atom typing and assignment of parameters and charges by analogy in a fully automated fashion, for the parametrization of modified RNAs that were not addressed in ref. [177] were enhanced; this allows for a more accurate parametrization of additional modified RNAs.[177] In this chapter, the latest advancements in modified RNA parametrization prompted the development a CHARMM[197]-based protocol to screen and investigate modified RNAs in complex with protein receptors. In this chapter, the methodology and applications of a computational protocol for the characterization of modified RNAs that can be recognized by a target protein, developed as part of this doctoral study is described. This protocol is a step toward filling the need for methods studying modified RNAs in the context of their interactions with proteins.

Methodology of the Protocol for the Characterization of Modified RNA : Protein Interactions

The protocol for the characterization of modified RNA : protein interactions can be outlined as follows: (1) a set of force field parameters for modified RNAs and a starting structure of an RNA : protein complex of interest are obtained for use as inputs to the protocol; (2) the modified RNAs under investigation are categorized into trees according to their structural and physicochemical properties to facilitate the screening, decision-making tool (3) a decision-making tool, comprising short simulations and interaction energy calculations, performs a fast and efficient search in a high-throughput fashion, through a list of different types of modified RNAs categorized into trees, and selects a subset of modified RNAs prone to interact with the target protein; (4) Selected modifications are further investigated using triplicate all-atom MD simulations and later evaluated and rated using association free energy calculations to produce a set of modifications expected to favor the interaction between an RNA strand and a given protein. The simulations also produce an ensemble of 3D structures of the modifications in complex with the protein of interest. These 3D structures can be analyzed to delineate the biophysical determinants of the interactions between selected modified RNAs and the target protein. The workflow of the protocol is shown in Figure III-1. In this section, the stages of the protocol for the characterization of modified RNA : protein interactions are described. A detailed description of the protocol and its use in a case study investigating the interactions of *E. coli* Polynucleotide Phosphorylase (PNPase) with a subset of modified RNAs is available in ref. [65].

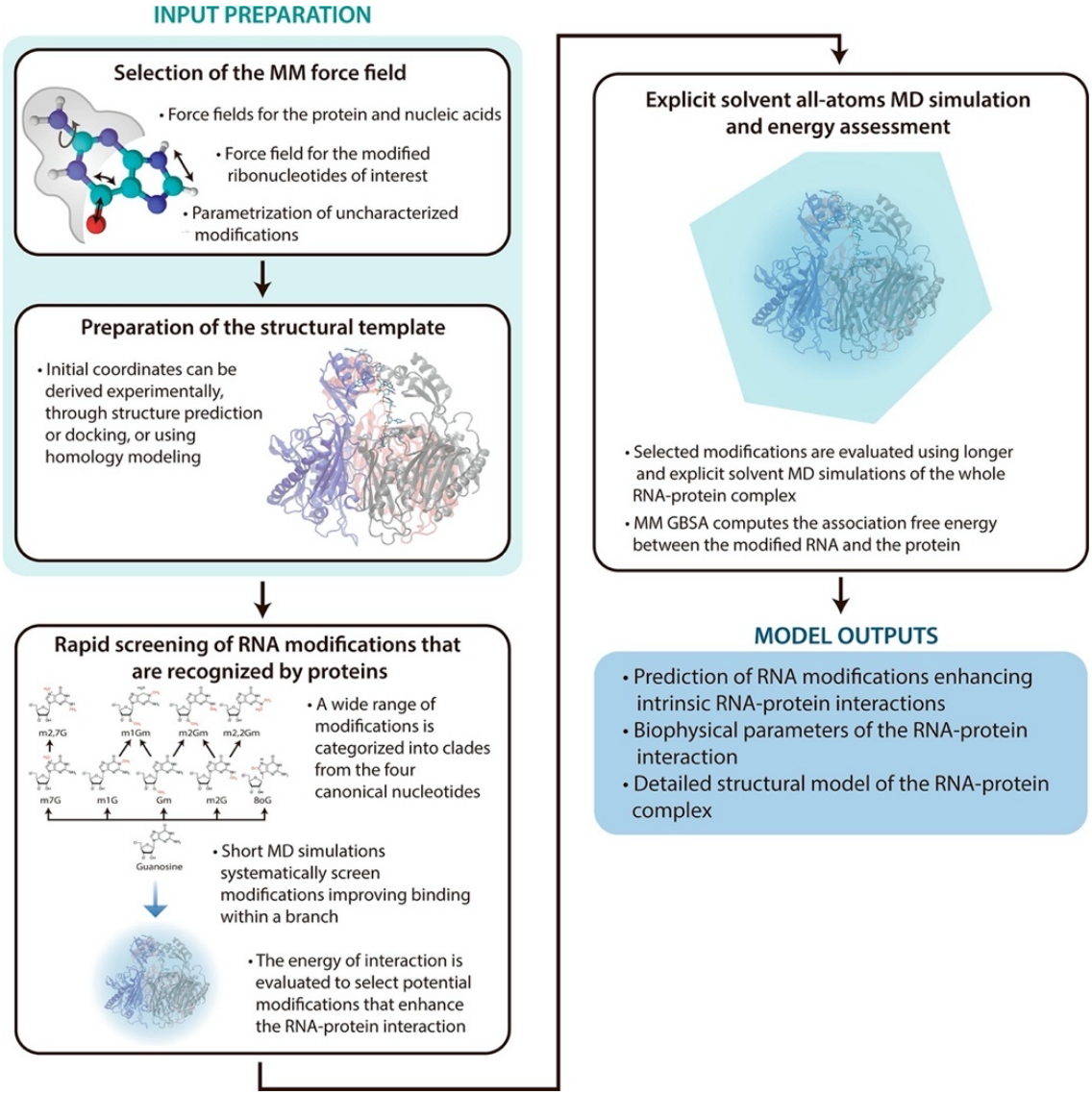

Figure III-1. Workflow for the protocol for the characterization of modified RNA : protein interactions. This figure is adapted from ref. [65*].

* Reprinted with permission from "A High-Throughput and Rapid Computational Method for Screening of RNA Post-Transcriptional Modifications that can be Recognized by Target Proteins" by Asuka A. Orr, Juan C. Gonzalez-Rivera, Mark Wilson, P. Reena Bhikha, Daiqi Wang, Lydia M. Contreras, and Phanourios Tamamis, 2018. *Methods*, 143, 34-47, Copyright 2018 by Elsevier.

Stage 1: Selection of Force-Field Parameters and Initial RNA : Protein Complex Structure

The central aspect of the protocol for the characterization of modified RNA : protein interactions is the use of MD simulations and energy calculations to identify modified RNAs that interact favorably with a target protein. An initial structure of the target protein in complex with an RNA strand and a molecular mechanics force field are necessary to describe the energetics underlying interactions of proteins and cognate modified RNA.

Several force fields are available for simulating proteins and nucleic acids, including CHARMM[42, 177, 198] and AMBER.[179, 199, 200] In the studies described in the next section, *Application of the Protocol for the Characterization of Modified RNA : Protein Interactions*, the CHARMM36 all-atom force field[42, 177, 198] is used to represent protein residues and nucleic acids. The recent parametrization of 112 naturally occurring modified RNAs[177] compatible with the CHARMM36 all-atom additive force field[42, 198] enabled the development and implementation of the protocol for the characterization of modified RNA : protein interactions. Further, with the release of the CHARMM parameters for modified RNAs,[177] the capabilities of CGenFF[193-196] to parametrize modified RNAs were improved.[177] Thus, additional modified RNAs that were not included in the parametrization of 112 modified RNAs[177] can be adequately parametrized using CGenFF with low penalties.[193-196] As part of this doctoral study, 13 additional modified RNAs were parametrized through CGenFF,[193-196] resulting in a total of 125 modified RNAs that can be interrogated for their interactions to target proteins. In the first version of the protocol

for the characterization of modified RNA : protein interactions, a subset of 46 modified RNAs were interrogated for their interaction with a target protein (*E. coli* PNPase);[65] in the second version of the protocol for the characterization of modified RNA : protein interactions, all 125 modified RNAs were interrogated for their interaction with target proteins (*E. coli* PNPase, human YTHDF1, human NOVA1, human TDP-43; currently under review).[67]

The initial structure of the target protein in complex with an RNA strand is used as a starting template to model the target protein interacting with modified RNAs. The initial structure can be (1) obtained from the PDB,[158] (2) built through homology modeling, or (3) built through first-principles or ab initio structure prediction techniques. The PDB[158] contains experimentally resolved protein structures, including RNA : protein complex structures. The public availability of these experimentally resolved structures can greatly facilitate computational studies; if the structure of an RNA : protein complex of interest has been resolved and deposited on the PDB,[158] the structure can be used as the initial structure. If the RNA : protein complex of interest is unavailable, but a homologous structure has been resolved, homology modeled can be an attractive method to build the complex of interest. Mutations can be introduced to the homologous complex to model the complex of interest using programs such as SCWRL4[201] with additional refinement through constrained minimizations and MD simulations.[202-204] If the RNA : protein complex of interest is not available and there is also no homologous complex structure available, then a combination of structure prediction and docking tools may be used to model the initial structure. The initial independent structures of a protein or RNA can be

obtained from existing experimentally resolved structures or can be modeled using a variety of structure prediction servers.[205-209] RNA docking programs can also subsequently be used to predict energetically favorable binding conformations of an RNA in complex with the target protein.[167, 168]

Stage 2: Categorization of Modified RNAs into Trees

The modified RNAs under investigation are categorized into "trees" according to their structural and physicochemical properties to facilitate the fast and efficient screening, decision-making tool (Stage 3). The modified RNAs are divided into four separate trees corresponding to one of the four canonical nucleotides (adenine, guanine, uracil, cytosine). Each tree starts from a seed comprising a canonical nucleotide and branches off into distinct modified RNAs (Figure III-2). The modified RNAs are categorized based on their structural and physicochemical properties such that modified RNAs within a branch all share similar properties. Modified RNAs stemming from a parent nucleotide corresponds to the parent nucleotide with the addition of simple chemical groups (Figure III-2). Thus, modified RNAs at the top of the trees are of higher complexity than the modified RNAs at the bottom of the trees (Figure III-2). The fast and efficient screening tool (Stage 3) will rapidly evaluate the modified RNAs energetic favorability for binding to a target protein based on the organization of the trees as described in the following section. The trees of modified RNAs used in the first version of the protocol for the characterization of modified RNA : protein interactions contained a subset of 46 modified RNAs and are presented in ref. [65]. The trees of modified RNAs used in the second version of the protocol

for the characterization of modified RNA : protein interactions contained all 125 modified RNAs[67] and are presented here in Figure III-2.

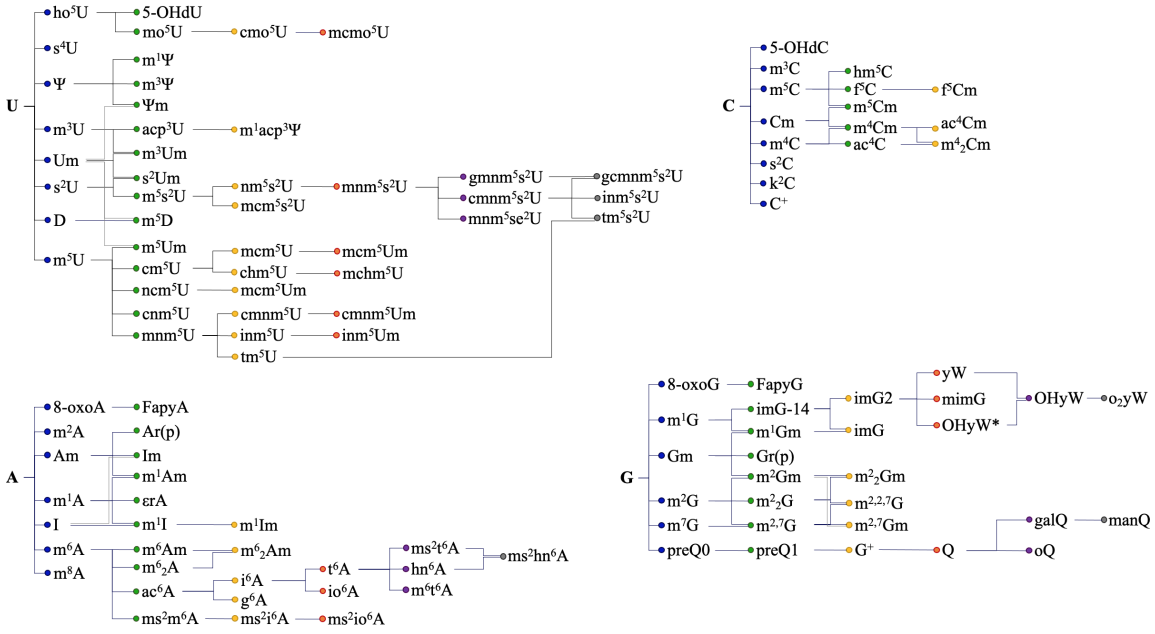

Figure III-2. Organization of modified RNAs into trees stemming from uracil (U), cytosine (C), adenosine (A), and guanosine (G).

Stage 3: Fast and Efficient Screening of Modified RNAs

The fast and efficient screening tool evaluates the modified RNAs in a high-throughput fashion to select a subset of modified RNA prone to interact with the target protein. The screening tool investigates the modified RNAs in accordance with their organization of the trees (Stage 2) using short MD simulations and energy calculations. For the purpose of the fast and efficient screening tool, the RNA : protein complex is truncated to reduce the computational resources required for the subsequent energy

minimizations and MD simulations. Light harmonic constraints are appropriately introduced to the truncated ends of the complex to preserve the shape and structure of the protein during the short MD simulations. Additionally, the GBMVII[210] implicit solvent model is used to quickly sample the nucleotides. Multi-ns MD simulations are later used with the entire, untruncated version of the RNA : protein complex as a final assessment (Stage 4) of the effect of the modified RNAs selected by the screening tool described in this section.

Following the organization of the trees, the fast and efficient screening tool first investigates the interactions of the canonical nucleotides through independent MD simulations and energy calculations. Each canonical nucleotide is introduced to a user defined RNA position under modification. Upon their introduction, short energy minimizations are performed to allow the four independent systems to adjust to the introduced canonical nucleotides. Subsequently, short MD simulations of 5 ns are performed for each of the four independent systems. Upon completion of the short MD simulations for the canonical nucleotides, the average interaction energy (the sum of electrostatic and van der Waals interaction energies) between the entire RNA strand and the truncated protein as well as the total energy of the RNA monomer strand is calculated for the last 1 ns of the short 5 ns MD simulations. These values are stored for subsequent comparisons to evaluate the favorability of the modified RNAs within the trees.

After the investigation of the canonical nucleotides corresponding to the seeds of each tree, the screening tool investigates modified RNAs. Starting from the first level of each tree, modified RNAs are independently introduced to the user defined RNA position

under modification. Upon their introduction, short energy minimizations followed by a short 2 ns MD simulations are performed. Subsequently, the average interaction energy between the entire RNA strand and the truncated protein as well as the total energy of the RNA monomer strand are calculated for the last 1 ns of the short MD simulation. The average interaction energy and total energy values are used to determine if the modified RNA is prone to interact favorably with the target protein. For the modified RNA to be considered prone to interact favorably with the target protein, the average interaction energy of the RNA strand with the modified RNA should be more favorable than the RNA strand containing the parent nucleotide and the total energy of the RNA monomer containing the modification must be approximately equal to or less than that of the native RNA. If these energetic criteria are met, then the modified RNA is selected and stored for further investigation. However, if the modified RNA does not meet these criteria, then the modified RNA is considered unfavorable and is screened out from further investigation along with the following levels of modified RNAs stemming from the unfavorable modified RNA. The fast and efficient screening tool continues to evaluate modified RNAs by levels of the trees until either the trees are fully investigated or all the remaining modified RNAs have been screened out or discarded.

The fast and efficient screening tool operates under the governing principle that further additions of simple chemical groups to a modified RNA with energetically unfavorable interactions are not expected to lead to any significant improvement in interactions. Thus, if a modified RNA is found to be unfavorable, then the modified RNAs in the branches stemming from the specific modified RNA are also discarded. The

governing principle was validated by investigating all modified RNAs in the fast and efficient screening tool in ref. [65].

Stage 4: Refinement and Final Evaluation of Selected Modified RNAs in Complex with the Target Protein

The selected modified RNAs from the fast and efficient screening tool (Stage 3) are further investigated using triplicate all-atom MD simulations and evaluated using association free energy calculations to produce a set of modified RNAs expected to favor the interaction between an RNA strand and a target protein. The selected modified RNAs are investigated using replicate, multi-ns, explicit solvent MD simulations to produce refined structures of the modified RNAs in complex with the target protein and to evaluate their energetic favorability. In this stage, the entire, untruncated structure of the RNA : protein complex is used as a starting template, and the selected modified RNAs are independently introduced at the user defined RNA positions. Triplicate simulation runs with different initial velocities are performed for reproducibility purposes.

Upon completion of the multi-ns simulations, the energetic favorability of the modified RNAs in complex with the target protein are assessed using MM–GBSA[49] association free energy calculations. The simulation snapshots serve as a statistical pool of conformations for the association free energy calculations as well as further energetic and structural analysis. Modified RNAs in complex with the target protein that acquire the most favorable association free energies that are more favorable than that of all canonical

nucleotides are considered to interact favorably with the target protein. These modified RNAs are considered as possible modifications that are recognized by the target protein.

Application of the Protocol for the Characterization of Modified RNA : Protein Interactions

The protocol for the characterization of modified RNA : protein interactions was developed for modified RNAs and successfully applied to several systems. The first version of the protocol (involving a subset of 46 modified RNAs) was used to study modified RNAs' interactions with *E. coli* PNPase.[65] The second version of the protocol (involving 125 modified RNAs) was used to study modified RNAs' interactions with human YTHDF1, human NOVA1, and human TDP-43.[67] Additionally, an analogous strategy to the protocol for the characterization of modified RNA : protein interactions was used to computationally design *E. coli* PNPase mutants with enhanced affinity to a single modified RNAs (8-oxoguanosine).[66] All experiments were performed by members of Dr. Contreras' lab at the University of Texas, Austin, TX. Contributors to studies associated with this section are listed in the authors lists of refs. [65, 66]. Additional details of these study are provided in ref. [65, 66].

Uncovering the Repertoire of Modified RNAs Recognized by *E. coli* PNPase

The protocol for the characterization of modified RNA : protein interactions was first applied to investigate modified RNAs binding to *E. coli* PNPase.[65] PNPase forms a homo-trimer complex that mainly functions as an exoribonuclease by dismantling RNA from the 3′ end to the 5′ end.[211, 212] While PNPase is not an essential protein in *E. coli*, it has been linked to a wide variety of cellular processes. For instance, PNPase has been implicated in the degradation of small RNAs, as well as damaged RNA in response to

oxidative stress by interacting directly with the RNA oxidative lesion 8-oxoguanosine (8-oxoG).[213, 214] Driven by PNPase's key role in RNA degradation,[215] in the first version of the protocol, a subset of all known modified RNAs, 46 modified RNAs out of approximately 150, were investigated to discern which modified RNAs are predicted to bind to *E. coli* PNPase.[65] The available X-ray structures of *E. coli* PNPase in complex with RNA fragments[216] and *Caulobacter crescentus* (*C. crescentus*) PNPase in complex with a 9-nucleotide RNA strand[217] were used in combination with the available CHARMM36-compatible all-atom additive force field parameters for 112 modified RNAs,[177] and the available CGenFF[193-196] parametrization for additional modified RNAs greatly facilitated this work.

In this application of the first version of the protocol for the characterization of modified RNA : protein interactions, the protocol for the characterization of modified RNA : protein interactions screened out ~89% of the total 46 possible number of modified RNAs to identify 5 modified RNAs with the most favorable interactions to *E. coli* PNPase.[65] Furthermore, the correlation between computationally derived MM–GBSA[49] free association energies and experimental EMSAs derived K_d values show reasonably high correlation (Figure III-3).[65] Thus, this test case showed that, using the protocol, (1) computational resources can be saved by only investigating, though all-atom MD simulations, RNA : protein complexes containing modified RNAs selected by the screening tool, and (2) experimental resources can also be reserved for only RNA strands containing the modified RNAs predicted to have the most favorable interactions to a protein.

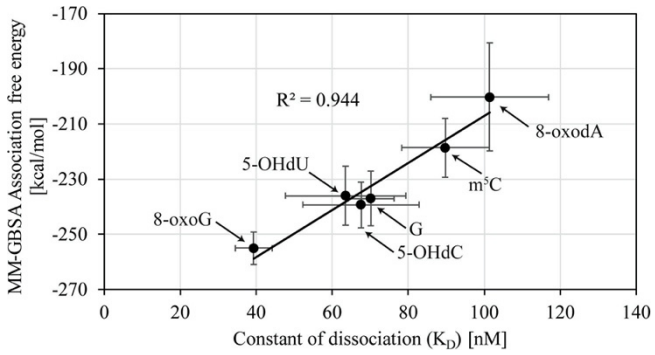

Figure III-3. Average MM–GBSA association free energies (kcal·mol^{-1}) with respect to experimentally derived dissociation constants (nM) of RNA strands containing select modified RNAs. The average and standard deviation MM–GBSA association free energy values for each modified RNA were calculated using the average association free energies calculated from the three independent simulation runs of each complex. This figure is adapted from ref. [65*].

Computational Evolution of PNPase Towards Enhanced Oxidized-RNA Binding

The protocol for the characterization of modified RNA : protein interactions was used as a basis to investigate an "inverse" problem of which amino acids enhance binding to a modified RNA of interest. Specifically, the analogous strategy was implemented to design PNPase mutants with enhanced affinity for oxidative RNAs (8-oxoG).[66] In this study, a library of possible amino acid mutations to *E. coli* PNPase was determined using a bioinformatics analysis, combinations of mutations were initially screened using short MD simulations in implicit solvent and interaction energy calculations of a truncated PNPase : RNA strand complex, and selected combinations of mutations were examined

* Reprinted with permission from "A High-Throughput and Rapid Computational Method for Screening of RNA Post-Transcriptional Modifications that can be Recognized by Target Proteins" by Asuka A. Orr, Juan C. Gonzalez-Rivera, Mark Wilson, P. Reena Bhikha, Daiqi Wang, Lydia M. Contreras, and Phanourios Tamamis, 2018. *Methods*, 143, 34-47, Copyright 2018 by Elsevier.

in detail using all-atom multi-ns MD simulations and association free energy calculations of the full complex system.[66]

The interactions of *E. coli* PNPase with 8-oxoG and G were first examined in atomic detail to provide insights into the mechanism of 8-oxoG discrimination. In this study, the PNPase subunits were hypothesized to cooperate to form a binding site using the dynamic SFF groove within the central channel of the PNPase homotrimer. Thus, this site was evolved using a novel approach combining biophysical constraints and a multi-stage computational screening strategy analogous to the protocol for the characterization of modified RNA : protein interactions to select the most promising mutants for enhanced 8-oxoG binding. In the multi-stage computational screening, PNPase mutants from a library of mutations, derived from bioinformatics analysis of the 782 PNPase sequences in the UNIPROT database,[218] analogous to ref. [219], were initially screened through short implicit solvent MD simulations and subsequently assessed using multi-nanosecond MD simulations and free energy calculations.

EMSAs performed by experimental collaborators showed that evolving just the SFF groove resulted in a fold increase in 8-oxoG affinity between 1.2 and 1.5 and/or selectivity between 1.5 and 1.9.[66] As a result of the computational analysis of the PNPase mutants, a few key observations were made: (1) mutations involving S76N and/or Y77F substitutions provide higher 8-oxoG affinity in the NYM, NYT and NFH mutants, attributed to new hydrogen bond interactions with 8-oxoG, (2) a unique trend seen in the GFT mutant is that the F78T substitution directly contacts the 8-oxoG modification, which can be linked to the increased 8-oxoG selectivity in this mutant, and (3), for the SYH

mutant, a balance between diminished 8-oxoG affinity at position P8 and increased 8-oxoG affinity at position P9 was observed.[66] These interactions are shown in Figure III-4. In addition to the improvement in 8-oxoG binding, complementation of K12 Δpnp with plasmids expressing mutant PNPases caused increased cell tolerance to H_2O_2.[66] This observation provides a clear link between molecular discrimination of RNA oxidation and cell survival. Moreover, this study provides a framework for the manipulation of modified RNA protein readers,[66] which has potential application in synthetic biology and epitranscriptomics.

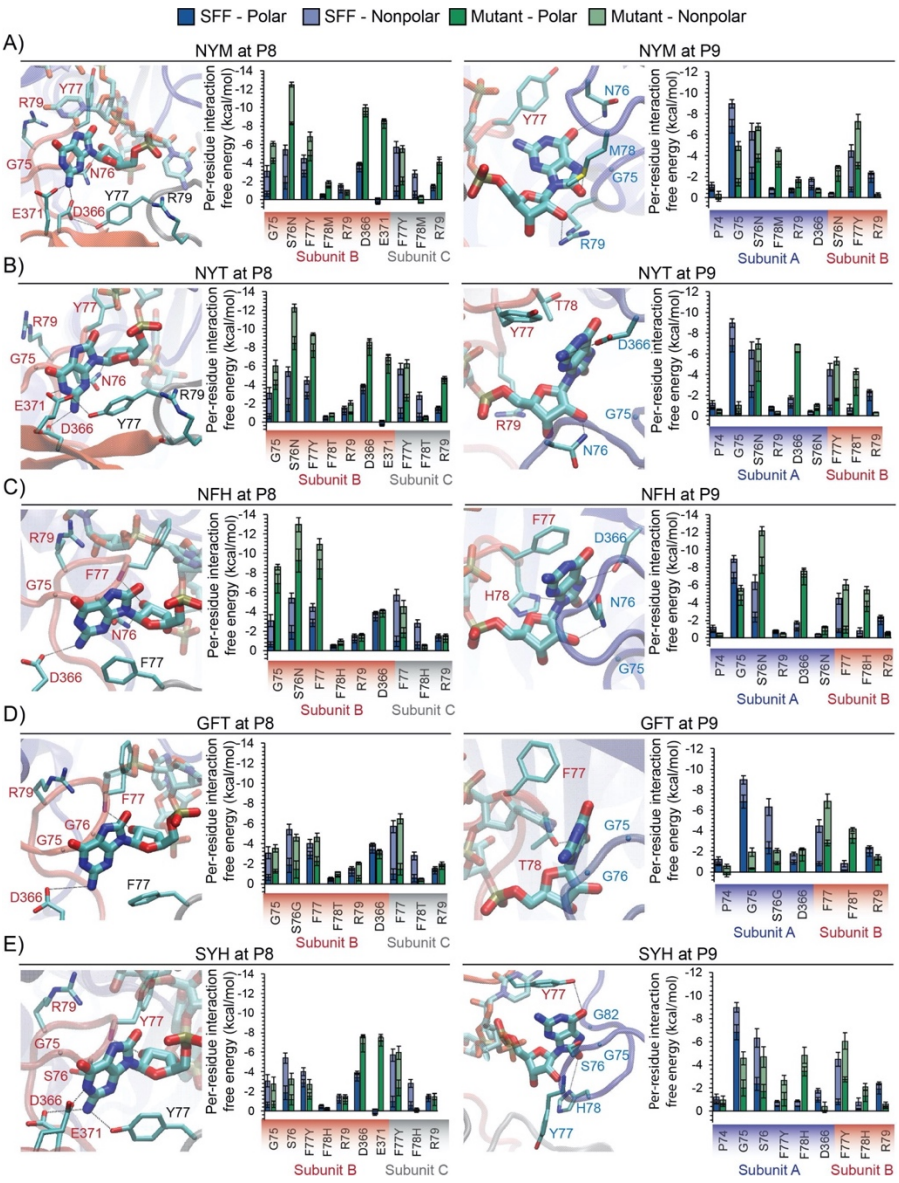

Figure III-4. Molecular interactions of 8-oxoG with the mutant PNPases: NYM (panel A for position P8 and panel B for position P9); NYT (panel C for position P8 and panel D for position P9); NFH (panel E for position P8 and panel F for position P9); GFT (panel G for position P8 and panel H for position P9) and SYH (panel I for position P8 and panel J for position P9). This figure is adapted from ref. [66*].

---

* Reprinted with permission from "Computational Evolution of an RNA-Binding Protein Towards Enhanced Oxidized-RNA Binding" by Juan C. Gonzalez-Rivera, Asuka A. Orr, Sean M. Engels, Joseph M. Jakubowski, Mark W. Sherman, Katherine N. O'Connor, Tomas Matteson, Brendan C. Woodcock, Lydia M. Contreras, and Phanourios Tamamis, 2020. *Computational and Structural Biotechnology Journal*, 18, 137-152, Copyright 2020 by Elsevier.

Uncovering a Broader Recognition of Modified RNAs Among Protein Readers

The protocol was expanded in its second version to include over 100 modified RNAs (125 modified RNAs) and applied for the first time in the literature to study modified RNA : protein interactions by four RNA binding proteins: *E. coli* PNPase, YTH N6-methyladenosine RNA binding protein 1 (YTHDF1), neuro-oncological ventral antigen 1 (NOVA1) and transactive response DNA-binding protein 43 (TDP-43).[67] These proteins were selected given that they each contain one of the three most representative currently known RNA binding domains (RBDs) in protein readers – the YT521-B homology (YTH) domain, the K homology (KH) domain, and the RNA recognition motif (RRM) domain. The 3D structure of *E. coli* PNPase in complex with an RNA strand was modeled in previous studies of this doctoral study,[65, 66] and the 3D structures of the remaining proteins in complex with RNAs have been experimentally resolved.[220-222] Moreover, all these domains are functionally characterized and/or implicated on human disorders.[223-231]

In this study, the hypothesis that protein readers exhibit an extensive (but previously undocumented) degree of promiscuity for recognition of modified RNAs was investigated. The protocol for the characterization of modified RNA : protein interactions demonstrates that, amongst the four biologically relevant proteins used as samples in this study, there is a broader range of modified RNAs recognized than initially anticipated.[67] Importantly, based on the protocol and experimental validation, some of the proteins were discovered to have a higher affinity for newly identified modified RNAs than to those modifications previously recognized and studied.[67] Specifically, *E. coli* PNPase was

computationally predicted and experimentally validated to recognize 1-methylguanosine (m^1G) and 8-oxoG;[67] YTHDF1 was computationally predicted and experimentally validated to recognize 3-methyluridine (m^3U) and 6-methyladenosine (m^6A);[67] NOVA1 was computationally predicted and experimentally validated to recognize 1-methylguanosine (m^1G) and 8-oxoG;[67] TDP-43 was computationally predicted and experimentally validated to recognize 1-methyladenosine (m^1A) and 6-methyladenosine (m^6A).[67] Overall, this study provides evidence of the possible broader binding promiscuity among protein readers of modified RNAs and provides insights into the molecular basis of their recognition. This study is currently under review.[67]

Concluding Remarks for the Protocol for the Characterization of Modified RNA : Protein Interactions

Chemical modifications to RNAs can impact the functions of RNA : protein interactions.[190] Uncovering which modified RNAs are recognized by proteins and how their binding affects the functions of RNA : protein interactions using conventional methods face challenges.[62-64] These challenges in uncovering which modified RNAs could be recognized by a target protein and the functional roles of these modified RNAs are compounded when considering that there are over 150 modified RNAs known to date.[60]

In this chapter, a protocol for the characterization of modified RNA : protein interactions was developed to uncover the repertoire of modified RNAs recognized by a target protein. The programs used in the protocol for the characterization of modified RNA : protein interactions has been successfully used to study an array of different RNA binding proteins. The programs have been expanded from the first version of the protocol[65] to include additional modified RNAs in the second version of the protocol,[67] thus 125 modified RNAs are accounted for in the current version. The application of the protocol for the characterization of modified RNA : protein interactions has uncovered the repertoire of modified RNAs recognized by protein readers and highlights the broader discrimination of these proteins than previously known.[65, 67] Additionally, an "inverse" problem of which amino acids enhance binding to a modified RNA of interest can also be approached using the strategies used in the protocol for the characterization of modified RNA : protein interactions.[66]

CHAPTER IV

INHIBITION OF AMYLOID SELF-ASSEMBLY

Study and Design of β-wrapin (Affibody Proteins) Binding and Specificity to Amyloidogenic Proteins

Introduction

Several age-related diseases are characterized by the accumulation of proteins into amyloid fibrils. Senile plaques formed by amyloid-β peptide (Aβ) and neurofibrillary tangles formed by tau are pathological features of Alzheimer's disease (AD);[232, 233] Lewy bodies and Lewy neurites formed predominantly by α-synuclein (α-syn) are pathological features of Parkinson's disease (PD) and the related illness, dementia with Lewy bodies;[234-238] Pancreatic islet amyloid formed by islet amyloid polypeptide (IAPP or amylin) are pathological features of type 2 diabetes (T2D).[239] The amyloid states formed by amyloidogenic proteins share similar self-assembly and structural cross β-spine properties, containing β-strands that are oriented perpendicular to the fibril axis.[240-249] Successful strategies to prevent protein aggregation and amyloid formation have been reported, which among others, include (i) inhibition of aggregation by sequestering individual amyloidogenic protein monomers, (ii) inhibition of aggregation through small molecules, and (iii) inhibition of amyloid growth through peptide-based inhibitors (reviewed in ref. [250]).

Inhibition of aggregation by sequestration is achieved through strong binding to aggregation-prone regions of amyloidogenic monomers by binding proteins, thereby

prohibiting self-assembly.[250] The sequestration of monomeric Aβ was achieved through an $ZA\beta_3$, an engineered Aβ-binding affibody protein homodimer with a disulfide bridge between the Cys28 residues of the two monomer subunits,[71, 72] hereinafter referred to as subunit 1 and subunit 2. $ZA\beta_3$ stabilizes a β-hairpin conformation of Aβ, which interacts with two β-strands formed by the affibody subunits to make a four-stranded intermolecular β-sheet[72] (Figure IV-1A). Thereby $ZA\beta_3$ prohibits the initial aggregation of Aβ monomers into toxic forms.[95] As tau, α-syn, and IAPP also contain motifs with high β-sheet propensities, β-wrapins (β-wrap proteins), derived from phage display libraries based on the $ZA\beta_3$ scaffold, were also discovered with the ability to recognize and stabilize β-hairpin conformations of motifs with high β-sheet propensity within the amyloidogenic proteins.[68-70] The introduction of between one and seven amino acid exchanges in $ZA\beta_3$ resulted in β-wrapins AS10, AS69, HI18, and TP4 which have different affinities for Aβ, α-syn, IAPP, and tau[68-73] (Figure IV-1B). According to experiments, AS10 binds to Aβ, α-syn, and IAPP with sub micromolar affinity,[73] AS69 binds significantly stronger to α-syn than to Aβ,[70] HI18 binds IAPP with nanomolar affinity,[69] TP4 binds significantly stronger to tau than to Aβ,[68] and $ZA\beta_3$ binds significantly stronger to Aβ than to α-syn or tau[68, 70, 72] (Figure IV-1B).

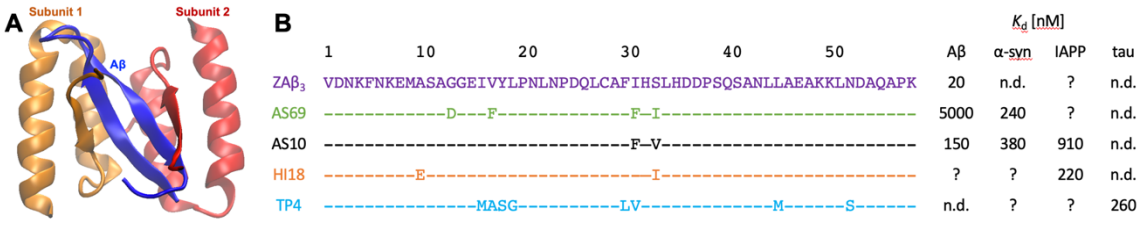

Figure IV-1. Overview of β-wrapin structure, sequence, and affinity to Aβ, α-syn, IAPP, and tau. (A) The structure of ZAβ$_3$ binding to Aβ. (B) Sequences of β-wrapin variants in comparison to ZAβ$_3$ and their affinities to Aβ, α-syn, IAPP, and tau as reported in refs. [68-73].

The development of β-wrapins with different affinities and specificities for Aβ, tau, α-syn, and IAPP could allow the discovery of multi-targeted binding agents that can serve in therapeutic applications for AD, PD, and T2D. Such agents can be of vital importance considering the increasing evidence for possible links between these proteins. For example in amyloid diseases, a significant portion of AD cases exhibit a third prevalent neuropathology, in addition to Aβ senile plaques and tau neurofibrillary tangles, associated with the aggregation of α-syn into Lewy bodies, which is traditionally associated with PD.[251, 252] Notably, patients with the Lewy body variant of AD exhibit an aggressive disease course with more pronounced cognitive dysfunction than patients with pure AD (reviewed in ref. [251]). Epidemiological studies also show an increased risk of AD in patients with T2D.[253-263] Plaques comprised of Aβ and IAPP co-aggregates were found in brain dissections from subjects diagnosed with both AD and T2D, and IAPP deposits were also found in subjects diagnosed with AD without clinically apparent T2D.[264] T2D is also associated with a significantly increased risk for developing PD.[265-268] Furthermore, clinical studies described that PD symptoms were notably worse after the onset of T2D

(reviewed in ref. [268]). Additionally, Aβ, tau, α-syn, and IAPP have been suggested to promote the formation and/or aggregation of each other.[269-273]

Single-targeted and multi-targeted β-wrapins with enhanced affinity for Aβ, α-syn, IAPP, and tau could serve as promising agents for AD, PD, and T2D. Experimental methods rely on the use of phage-display libraries to discover novel β-wrapins with mutable β-wrapin positions selected based on intuition and insights from NMR studies. Despite the fact that phage-display methods are significantly more expensive compared to computational design methods, they have been successfully implemented to derive a β-wrapin variant (ZSYM73) with pico-molar affinity to Aβ.[21] However, two challenging aspects of β-wrapin design are difficult to resolve through existing tools alone and still remain challenging: (1) the identification of β-wrapin residue positions that, upon mutation, are most likely to result in higher-affinity and (2) the identification of mutations to these β-wrapin residue positions that can be introduced to design β-wrapin variants with enhanced single-targeted or multi-targeted affinity toward amyloidogenic proteins.

In this chapter, the methodologies developed to study and design β-wrapins binding to amyloidogenic proteins as part of this doctoral study and their applications is described. The methodology to design β-wrapins with enhanced affinity to Aβ, tau, α-syn, and IAPP or combinations of the monomers involves an iterative, systems-like approach involving the study of β-wrapins binding to amyloidogenic proteins with known, experimentally derived affinities through a combinations of MD simulations and free energy calculations, an optimization-based protein design model that performs an initial screening, MD simulations of the designed β-wrapins passing the initial screening, and

free energy calculations evaluating the designed β-wrapins. Additionally, the rational design of β-wrapins with aromatic canonical and non-canonical amino acid N-termini modifications mimicking the mechanism of action of a particular compound binding to amyloid fibrils and inhibiting amyloid elongation is also explored using MD simulation, structural analysis, and free energy calculations in this chapter. The rational design of modified β-wrapin N-termini was performed with the aim to amplify β-wrapins' amyloid inhibitory properties and overcome difficulties in designing β-wrapins with significantly enhanced multi-targeted properties.

Methodology for the Study and Design of β-wrapins Binding to Amyloidogenic Proteins

To design β-wrapins with enhanced single-targeted affinity to Aβ, tau, α-syn, and IAPP or multi-targeted affinity to combinations of the amyloidogenic protein monomers, an iterative, systems-like approach was used following three steps (Figure IV-2): (1) understand the binding and specificity of β-wrapins for amyloidogenic proteins, (2) screen designed β-wrapin sequences using an in-house developed optimization-based design model, and (3) evaluate the top-ranked designed β-wrapin sequences using all-atom MD simulations and free energy calculations.

To understand the binding and specificity of β-wrapins for amyloidogenic proteins, molecular dynamics (MD) simulations and free energy calculations were performed to investigate β-wrapins in complex with Aβ, α-syn, IAPP, and tau. For this stage, the NMR structures of β-wrapins $ZA\beta_3$, AS69, and HI18 in complex with Aβ,[71] α-syn,[70] and IAPP[69] respectively, provided the grounds to computationally investigate a series of β-wrapins in complex with amyloidogenic monomers. The investigated β-wrapins initially comprised of previously published β-wrapins whose affinities to Aβ, α-syn, IAPP, and/or tau have been experimentally evaluated in previous studies.[68-73]

Biophysical insights gained from these studies were passed to the screening, optimization-based design model as biophysical constraints in the form of β-wrapin residue positions amenable for modification. To initial screen designed sequences of β-wrapins for enhanced single-targeted affinity to Aβ, tau, α-syn, and IAPP or multi-targeted combinations of the amyloidogenic protein monomers, a FORTRAN design program was developed. The design program introduces mutations to a dynamic structural template

from MD simulations and evaluates their energetic favorability in a coarse-grained fashion, treating each residue as a single bead, allowing for a rapid initial screening of potential designed β-wrapin variants.

To evaluate the top-ranked designed sequences of β-wrapin variants produced from the optimization-based model, the designed β-wrapins are modeled and evaluated through MD simulations, free energy calculations, and structural analysis. Designed β-wrapin variants with association free energies less than the most effective existing β-wrapins and meeting criteria for residue pair-wise interactions and structural stability in the unbound state (data gathered from the studies investigating the binding and specificity of existing β-wrapins, described in the following paragraphs) can be considered promising designs. Data from MD simulations and free energy calculations for all the evaluated designs were additionally passed back to the first stage investigating the binding and specificity of β-wrapins to determine any new "rules" (e.g. residue positions that should not be mutated as mutations can result in structural instability) to be applied to the screening of designed β-wrapin sequences in the optimization-based design model.

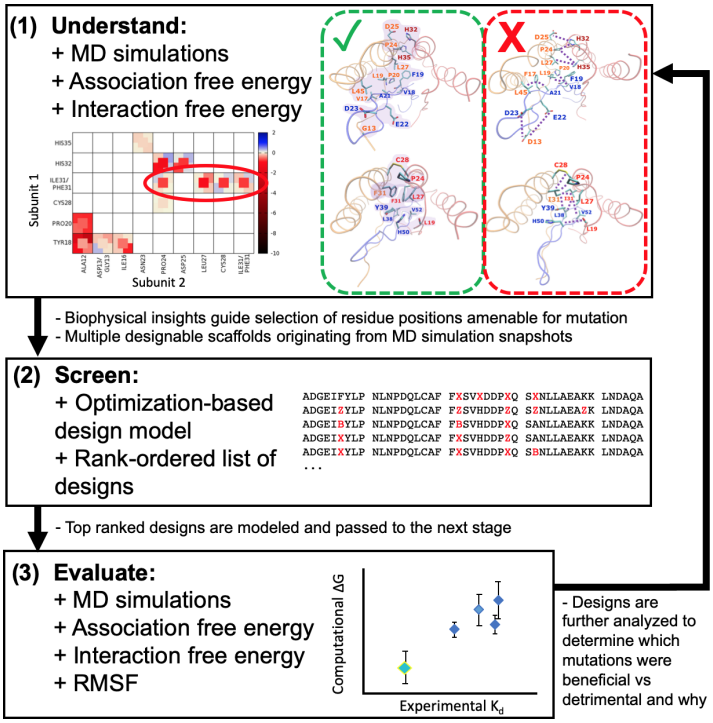

Figure IV-2. Workflow for the β-wrapin design strategy. (1) The binding and specificity of experimentally tested β-wrapins are understood using MD simulations, association free energy calculations, and residue-pairwise interaction free energy calculations. (2) Designed β-wrapin sequences are initially screened using an optimization-based design model, which produced a rank-ordered list of designs. (3) The top-ranked designed β-wrapin sequences are modeled and evaluated using MD simulations, association free energy calculations, residue-pairwise interaction free energy calculations, and RMSF calculations of the β-wrapin's unbound state.

Stage 1: Understanding the Binding and Specificity of β-wrapins for Amyloidogenic Proteins

To design β-wrapins with enhanced binding to amyloidogenic proteins, the binding and specificity of existing β-wrapins must be examined and understood. To understand the binding and specificity of existing β-wrapins, MD simulations and free energy calculations are used. In this stage, the NMR structures of the ZAβ3 : Aβ,[71] AS69 : α-

syn,[70] and HI18 : IAPP[69] complex structures provide initial structural templates to model different β-wrapin variants in complex with different amyloidogenic proteins. The mutated side chains across different β-wrapins are introduced through SCWRL4.0,[201] and the Cys28-Cys28 disulfide bridge between the two subunits of each β-wrapin was maintained using a disulfide patch from the CHARMM36 topology file.[198] Additionally, truncated N- and C-termini of the modeled β-wrapins were acetylated and amidated to alleviate any artifacts from artificially placing positively or negatively charged groups at the backbone termini of the truncated complexes.

The modeled β-wrapins in complex with the amyloidogenic protein monomer of interest were subsequently investigated through MD simulations. Prior to the execution of the MD simulations, each complex was solvated in a cubic explicit-water box with a potassium chloride concentration set to 0.15 M. Additional potassium or chloride ions were introduced to neutralize the charge of the simulation system. The ions were introduced to the simulation through Monte Carlo simulations.[274, 275] The simulation systems were subsequently energetically minimized through steepest descent minimization and Adopted Basis Newton–Raphson minimization.

The simulation systems were equilibrated in three steps, analogously to ref. [276]. In the first equilibration step, each complex was simulated for 1.0 ns with 1.0 $kcal \cdot mol^{-1} \cdot Å^{-2}$ harmonic constraints on backbone atoms and 0.1 $kcal \cdot mol^{-1} \cdot Å^{-2}$ harmonic constraints on side chain atoms. In the second equilibration step, all constraints were released, and the systems were simulated for 10 ns to allow the modeled and mutated residues of the β-wrapins to adopt improved conformations and interactions. Finally, an additional six

equilibration runs were performed, according to which the refined-equilibrated structure of the second step was extracted and subjected to an additional short simulation with light constraints to allow the simulated systems to further equilibrate. In this step, the velocities were reinitialized to produce six independent runs for each complex.

Finally, all constraints were released, and the simulation systems entered the final production runs in which each simulation system was simulated for 24 ns with frames extracted every 20 ps. Six independent runs for each β-wrapin : amyloidogenic protein complex system were performed to check the reproducibility of results. All simulations were performed using the Leap-Frog Verlet algorithm with the pressure set to 1.0 atm and the temperature held at 300 K using the Hoover thermostat. Fast table lookup routines for nonbonded interactions[277] were applied, and the SHAKE algorithm was implemented to constrain the bond lengths to hydrogen atoms.[278, 279] Periodic boundary conditions were applied in each simulation, and all MD simulations were performed using CHARMM36 topology and parameters.[198] The simulations were performed using CHARMM.[197] Upon the completion of the MD simulation runs, water molecules and ions were stripped from the trajectories, and the sextuplicate 24 ns simulation snapshots were analyzed as follows.

The binding of the β-wrapin variants to the amyloidogenic protein monomers were then evaluated through free energy calculations. To uncover the energetic driving forces leading to high affinity and no/minimal affinity of β-wrapins for amyloidogenic protein monomers, MM–GBSA[49] association free energy calculations were performed. To uncover the key interactions leading to enhanced or diminished affinity of β-wrapins for

amyloidogenic protein monomers residue pairwise interaction free energies were performed, analogously to refs. [202, 276, 280].

Stage 2: Optimization-Based Design of Promising β-wrapin Variants

Based on the biophysical insights gained from the previous stage, a protein-protein design program was devised, inspired by Protein WISDOM,[219] to computationally identify the set of optimum set of positions and mutations which will increase the binding affinity of the designed β-wrapin to a target amyloidogenic protein without interfering the designed β-wrapin's conformation ((2) of Figure IV-2). The program uses biophysical insights from MD simulations ((1) of Figure IV-2) in the form of residue pairwise interaction free energy values as input to determine residue positions amenable for mutation and exhaustively screens a library of possible mutations per mutable residue position ((2) of Figure IV-2). Additionally, structures snapshots extracted from MD simulations were used as structural templates on which mutations were introduced. Upon completion of the screening stage, combinations of mutations were rank ordered by their interaction energy from most favorable to least favorable. The optimization-based problem is presented in Equation IV-1 and is solved through exhaustive enumeration in FORTRAN.

$$\min\left[\sum_{U=1}^{p}\frac{1}{f}\sum_{s=1}^{f}\left(\sum_{i=n+1}^{n+2b}\sum_{j_i=1}^{m_i}\sum_{\substack{k=1\\k\neq i}}^{n+2b}\sum_{l_k=1}^{m_k}\left(E_{ik}^{j_i l_k}\cdot w_{ik}^{j_i l_k}(s)\cdot y_i^{j_i}\cdot y_k^{l_k}\cdot a_{ik}^{j_i l_k}\right)+\gamma\cdot\sum_{i=n+1}^{n+2b}\sum_{j_i=1}^{m_i}\left(\mathrm{SASA}_i^{j_i}(s)\cdot y_i^{j_i}\right)\right)\right] \quad \text{Equation IV-1}$$

Subject to

$$\sum_{j_i=1}^{m_i} y_i^{j_i}=1\forall i \quad \text{Equation IV-2}$$

$$y_i^{j_i},y_k^{l_k},w_{ik}^{j_i l_k}(s)\in\{0,1\}\forall i,j_i,k,l_k$$

$$a_{ik}^{j_i l_k}\in\{0.5,1\}\forall i,j_i,k,l_k$$

$$li\in\{\varphi_1,\varphi_2,\ldots,\varphi_v\}$$

$$j_{i+b}=j_i\forall(n<i\leq n+b)$$

$$j_{i-b}=j_i\forall(n+b<i\leq n+2b)$$

$$l_k=j_i\forall k\in\{\varphi_1,\varphi_2,\ldots,\varphi_v\}$$

$$lk>i\forall k\in\{\varphi_1,\varphi_2,\ldots,\varphi_v\}$$

In Equation IV-1 and Equation IV-2, p is the total number of amyloidogenic proteins under investigation, and U is the an index running through each of the amyloidogenic proteins under investigation. f is the total number of flexible structural templates under consideration in the design (10; extracted from MD simulations performed in the previous stage[74, 75]), and s is an index running through each flexible structural template. n is the total number of residues in the targeted amyloidogenic protein, and b is the total number of residues per β-wrapin subunit. i is a residue position in the β-wrapin modified. j_i is one of a set of possible residues, m_i, for position i. k is one of the residue position interacting with the residue of position i, with $k \neq i$. l_k = one of 20 possible residues, m_k, for position

k. $E_{ik}^{j_i l_k}$ represents the interaction energy between residue j_i at position i and residue l_k at position k, and the values were extracted from the SIPPER force field[234]. In line with the coarse-grained representation of residues in the SIPPER force field[234], $w_{ik}^{j_i l_k}(s)$ is binary variable equal to 1 if the geometric centers of residue l of position k is within 6.0 Å of residue j of position i and 0 otherwise at the specific snapshot s. $y_i^{j_i}$ and $y_k^{l_k}$ were binary variables equal to 1 if positions i,k were occupied by residues j_i,l_k and 0 otherwise. $a_{ik}^{j_i l_k}$ is a binary variable equal to 0.5 if $n < i \leq n+b$ and $n < k \leq n+b$ or if $n+b < i \leq n+2b$ and $n+b < k \leq n+2b$ and equal to 1 otherwise; in this way, intramolecular interactions within a subunit are balanced with the rest of the interactions and are not counted twice; notably intramolecular interactions are explicitly considered in the design. $\{\varphi_1,\varphi_2,\ldots,\varphi_v\}$ is the set of residue positions amenable for modification in the β-wrapin such that $\varphi_1<\varphi_2<\ldots<\varphi_v$. $\mathrm{SASA}_i^{j_i}(s)$ is the estimated solvent accessible surface area (SASA) of the introduced amino acid j_i at position i and is estimated as the theoretical SASA of the original residue – the SASA of the original residue in the scaffold at frame s). γ is an empirical surface tension coefficient (0.001 kcal·mol^{-1}·Å^{-2}). The constraints $j_{i+b} = j_i \forall (n<i\leq n+b)$ and $j_{i-b} = j_i \forall (n+b<i\leq n+2b)$ guarantee that the two β-wrapin subunits will be identical upon mutation. The aforementioned constraint is optional and may not be applied. $l_k=j_i \forall k \in \{\varphi_1,\varphi_2,\ldots,\varphi_v\}$ is a constraint to ensure that the residue at position k must be the same as the residue at position i for k belonging to the set of amenable positions, and $lk>i \forall k \in \{\varphi_1,\varphi_2,\ldots,\varphi_v\}$ is a constraint to ensure unique pair-wise interactions. The

computational design algorithm reads the coordinates of the amyloidogenic proteins first, followed by the coordinates of the two β-wrapin subunits (hence the summation of *i* begins at *n*+1).

Upon solution of the design model using exhaustive enumeration in FORTRAN, the generated β-wrapin variant sequences were ranked by interaction energy from lowest (most favorable) to highest (least favorable). The top-ranked designed β-wrapin variant sequences were subsequently selected to be evaluated based on their energetic favorability in binding a target amyloidogenic protein or set of amyloidogenic proteins, the presence of enhanced or diminished interactions, and structural stability in their unbound state ((3) of Figure IV-2).

Stage 3: Evaluation of Top Ranked β-wrapin Variants

To evaluate their energetic favorability, the designed β-wrapin variants were modeled and simulated in triplicate, explicit solvent MD simulations, and using snapshots from the MD simulations, association free energy calculations were performed to evaluate the designed β-wrapin variants' energetic favorability. For the designed β-wrapin variants with predicted enhanced affinity (based on association free energy calculations), three additional replicate MD simulations were performed for reproducibility for a total of six replicate MD simulations per promising design. To identify enhanced interactions leading to enhanced predicted affinity and ensure no interaction-based switches leading to diminished affinity were present,[74, 75] interaction free energy calculations were performed. Additionally, to account for the entropic component of the β-wrapins' affinity to the target

amyloidogenic protein, an additional criterion was introduced based on MD simulations of the unbound state of the designed β-wrapins investigating their stability in comparison to existing β-wrapins. To evaluate the β-wrapin's stability, MD simulations of the unbound β-wrapins in explicit water and root mean square fluctuation (RMSF) calculations on the resulting MD simulation snapshots were used. The RMSF of the Cα atoms for residues 21 through 56 were calculated. The Cα atoms for residues 12 through 20 were excluded in these calculations as this region is disordered in the unbound state of the β- wrapins. Higher fluctuations in the designed β-wrapins compared to existing β-wrapins, resulting in higher average RMSF values, were hypothesized to correlate with more disorder or higher entropy in the unbound state of the designed β-wrapins compared to existing β-wrapins. The average RMSF was found to be correlated with the entropy for experimentally tested β-wrapins (according to experiments performed by Dr. Hoyer's lab). Thus, the average RMSF was used as a metric to evaluate the designed β-wrapin variants' entropic favorability.

Applications for the Study and Design of β-wrapins Binding to Amyloidogenic Proteins

The methodology to study and design β-wrapins has been applied to study how β-wrapin variants bind to amyloidogenic proteins Aβ, α-syn, and IAPP[74, 75] as well as their binding to variants of tau. All experiments in this section were performed by members of Dr. Hoyer's lab at Heinrich-Heine-University, Düsseldorf, Germany. Contributors to these works are listed in the authors lists of refs. [74, 75] and will be listed as authors in additional papers to be submitted. Additional details of these studies are provided in refs. [74, 75].

Study of β-wrapin variants binding Aβ, α-syn, and IAPP

To understand ((1) of Figure IV-2) the binding properties of β-wrapin variants in complex with Aβ, α-syn, and IAPP were investigated through a combination of computational and experimental methods.[74, 75] The β-wrapin variants binding to Aβ, α-syn, and IAPP were investigated as they have a range of affinities for all three amyloidogenic proteins, ranging from micro-molar affinities to undetectable binding, and NMR structures have been resolved depicting β-wrapin binding to the three amyloidogenic proteins; structures of any β-wrapin in complex with tau have not been reported.

According to MD simulations and free energy calculations, β-wrapin complexes acquiring significantly favorable polar association free energies in conjunction with fairly favorable nonpolar association free energies are indicative of a highly active β-wrapin.[74, 75] Additionally, Aβ $_{18}$VFFAED$_{23}$ and α-syn $_{38}$LYVGSK$_{43}$ are key residue domains determining a β-wrapin's binding specificity, whereas IAPP is a more promiscuous β-

wrapin target, presumably due to the lower number of charged residues in its β-hairpin motif compared to Aβ and α-syn.[74, 75] Due to the lower selectivity of IAPP compared to Aβ and α-syn in binding to β-wrapins, $ZA\beta_3$ and AS69, which were engineered to bind to Aβ and α-syn, respectively, were computationally predicted and experimentally verified to have similar affinities to IAPP as AS10.[74] Finally, according to the analysis, common non-polar and polar interactions between β-wrapin residues and corresponding amyloidogenic protein (Aβ, α-syn, and IAPP) residues, based on structural alignment, contribute to AS10's ability to bind all three amyloidogenic proteins with micro-molar affinity (Figure IV-3).[74, 75] These studies suggested that the design of high-affinity single-targeted or multi- targeted β-wrapin variants could be achieved through (1) optimization of specific interactions with corresponding target residues in the complex core that forms upon coupled folding-binding, (2) exploiting the dynamic interactions with peripheral segments of the amyloidogenic proteins that remain structurally flexible in the bound state.[74, 75]

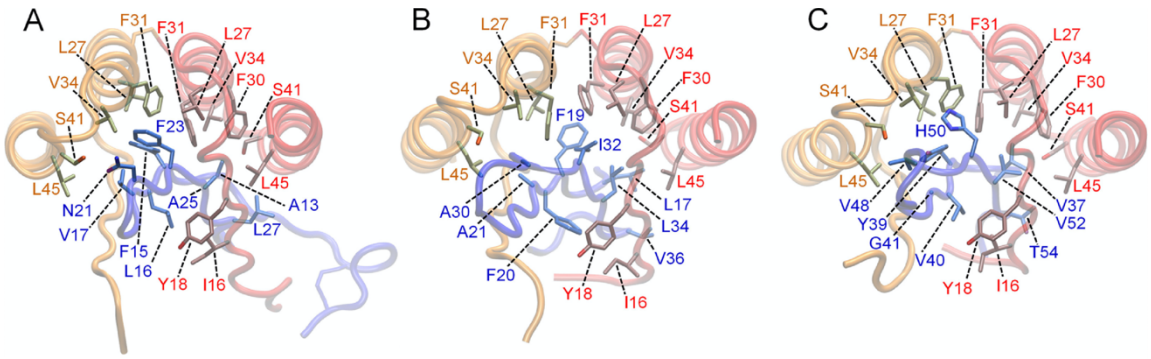

Figure IV-3. Molecular graphics images of the common hydrophobic interactions of AS10 in complex with (A) IAPP, (B) Aβ, and (C) α-syn. AS10 subunits 1 and 2 are shown in red and orange tube representation, respectively, and IAPP is shown in blue tube representation. The specified hydrophobic and aromatic interactions contribute significantly to the ability of AS10 to sequester all three of the amyloidogenic proteins. This figure is adapted from ref. [74*].

Molecular modeling of the TP4 : tau complex

To understand ((1) of Figure IV-2) the binding properties of β-wrapin variants in complex with tau and tau variants were also computationally investigated. Due to the absence of any experimentally resolved structure of a β-wrapin : tau complex, the binding of TP4 in complex with tau was computationally modeled. NMR spectroscopy of TP4 in complex with tau identified two alternative tau binding sites comprising tau domains 267-317 and 297-358.[68] Using the NMR structure of ZAβ$_3$ in complex with Aβ^{72} as a template, a combination of homology modeling and simulation techniques, including MODELLER,[206] constrained REMD simulations, and extensive MD based refinement

* Reprinted with permission from "Elucidating the multi-targeted anti-amyloid activity and enhanced islet amyloid polypeptide binding of β-wrapins" by Asuka A. Orr, Hamed Shaykhalishahi, Ewa A. Mirecka, Sai Vamshi R. Jonnalagadda, Wolfgang Hoyer, and Phanourios Tamamis, 2018. *Computers & Chemical Engineering*, 116, 322-332, Copyright 2018 by Elsevier.

were used to model missing residues of tau loop regions and generate model structures of TP4 in complex with the two binding sites of tau, independently (Figure IV-4B).

Following the modeling of TP4 in complex with the two binding sites of tau, the accuracy of the initial structure was evaluated through MD simulations and free energy calculations of TP4 in complex with tau variants with experimentally determined affinities for TP4.[68] According to sextuplicate MD simulations and association free energy calculations, as described in *Methodology for the Study and Design of β-wrapins to Amyloidogenic Proteins* and performed in previous studies,[74, 75] high agreement between the computationally derived association free energies and previously derived experimentally derived K_d values[68] was achieved for both binding sites of tau (Figure IV-4A). This supported the accuracy of the modeled TP4 : tau complex structures. Additionally, the computationally derived association free energies indicated that TP4 binding to tau binding site 1 (residue domain 267-317) was more energetically favorable than tau binding at site 2 (residue domain 297-358). Thus, dual-Aβ/tau β-wrapin design studies described in the following passages focused on Aβ and tau binding side 1.

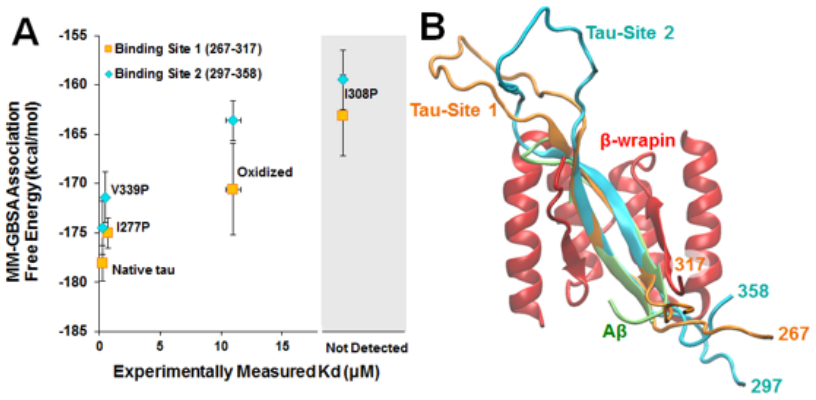

Figure IV-4. Computational modeling of TP4 in complex with tau. (A) High correlation between the computationally predicted association free energies calculated using the MM-GBSA approximation and previously derived experimentally measured affinities[68] of the TP4 binding to tau binding site 1 and 2. (B) TP4 : tau (binding sites 1 and 2) in comparison with $ZA\beta_3$: Aβ complex structure.

Design of single-targeted β-wrapin variants binding IAPP with enhanced affinities

The design program was applied in several rounds to screen and generate β-wrapin variants with enhanced affinity to IAPP ((2) of Figure IV-2). In the design program, HI18 was used as a basis as it has the highest affinity to IAPP of all the existing studied β-wrapin variants[69] (Figure IV-1). MD simulations combined with association free energy calculations and residue-pairwise interaction free energy calculations were used to evaluate the affinity or enthalpic favorability of the designed β-wrapin variants passing the initial screening by the design program ((3) of Figure IV-2). MD simulations combined with RMSF calculations were used to evaluate the stability or entropic favorability of the designed β-wrapin variants ((3) of Figure IV-2).

Based on the design program, MD simulations and free energy calculations of the top-ranked designed β-wrapins in complex with IAPP, and MD simulations and RMSF

calculations of the unbound designed β-wrapins, two promising, asymmetrically designed β-wrapins, named HI18-R5-3 and HI18-R5-6 (Table IV-1), were predicted to have enhanced affinity to IAPP compared to HI18. Both designed β-wrapins originated from the asymmetric design round in which HI18 subunit 1 residue positions I16, P20, I34, S41, and A42 as well as subunit 2 residue positions G11, A12, S39, and S41 were selected as positions amenable for mutation. The particular residue positions were selected aiming to target charged or polar residues of IAPP, according to the biophysical insights derived in ref. [74]. The targeting of charged or polar residues of IAPP aimed to significantly improve polar association free energies as previous studies on β-wrapin : amyloidogenic protein complexes, described in the previous sections, suggest that a low polar association free energy in addition to a fairly low non-polar association free energy is characteristic of β-wrapins with high affinity.[74, 75]

Table IV-1. Summary of the most promising designed β-wrapin variants for IAPP. ΔΔG values are calculated as the association free energy of the designed β-wrapin variant minus the association free energy of HI18. All statistics are averaged over six individual simulation runs.

β-wrapin Variant	**Subunit 1 Mutations to HI18**	**Subunit 2 Mutations to HI18**	**ΔΔG (kcal·mol^{-1})**	**RMSF (Å)**
HI18 (wild type)			0.0 ± 5.2	2.6 ± 0.1
HI18-R5-3	P20A, I34V, S41N, L45V	A12Q, S39K, A42D	-19.3 ± 2.2	1.8 ± 0.1
HI18-R5-6	I34W, S41N, L45V	S11Q, A12Q, S39K, A42D	-15.4 ± 3.2	2.2 ± 0.1

The mutations (introduced to the sequence of HI18), energetic favorability (based on ΔΔG compared to HI18), and stability (based on average RMSF) of the two designed β-wrapins are summarized in Table IV-1. In the following text, the predicted improved interactions of HI18-R5-3 and HI18-R5-6 compared to HI18, based on interaction free energy calculations and visual inspection of the six replicate simulation trajectories, are described.

Improved interactions of HI18-R5-3: The predicted enhanced affinity of designed β-wrapin, HI18-R5-3, to IAPP compared to HI18 can be attributed to mutations both directly and indirectly allowing or enhancing hydrogen bonds to IAPP within the simulations. In subunit 1 (Table IV-1, second column), the P20A and S41N mutations directly enhance hydrogen bonds to IAPP, and the I34V and L45V mutations indirectly enhance hydrogen bonds to IAPP by decreasing congestion within the complex structure (Figure IV-5). Specifically, the P20A allows for the formation of a hydrogen bond to N14 of IAPP (Figure IV-5). This interaction also greatly enhances the β-sheet interaction between L19 of subunit 1 and F15 of IAPP (Figure IV-5). Additionally, the S41N mutation enhances the hydrogen bond to N21 of IAPP (Figure IV-5). Both the I34V and L45V mutations decrease congestion to enhance the hydrogen bond between β-wrapin position 41 and N21 of IAPP (Figure IV-5).

In subunit 2, (Table IV-1, third column), the A12Q and A42D mutations directly allow for hydrogen bonds or salt-bridges to IAPP (Figure IV-5), and the S39K mutations indirectly enhances HI18-R5-3's affinity to IAPP by stabilizing the unbound state of the designed β-wrapin. The A12Q mutation allows for the formation of a hydrogen bond to

T30 of IAPP (Figure IV-5). This interaction also enhances the nonpolar interaction between the N-terminus of subunit 2 and the C-terminus of IAPP (Figure IV-5). Additionally, the A42D mutation allows for the formation of a salt-bridge with K1 of IAPP. The S39K mutation supports the A42D mutation and stabilizes the isolated designed β-wrapin by forming a salt-bridge with A42D (not shown); this is crucial as previous, preliminary studies indicated that a lone A42D mutation would destabilize the unbound β-wrapin. The presence of the S39K mutation in conjunction with the A42D mutation proved beneficial in the simulations as the average RMSF of the isolated HI18-R5-3 is not elevated compared to the isolated HI18 (Table IV-1, fifth column).

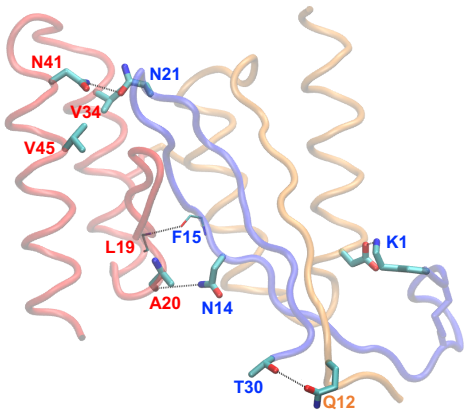

Figure IV-5. Molecular graphics image showing the improved interactions of HI18-R5-3 in complex with IAPP. IAPP is shown in blue tube representation, subunit 1 of HI18-R5-3 is shown in red tube representation, and subunit 2 of HI18-R5-3 is shown in orange tube representation. Blue labels indicate residues of IAPP. Red and orange labels indicate residues of subunits 1 and 2 of HI18-R5-3, respectively.

Improved interactions of HI18-R5-6: The predicted enhanced affinity of designed β-wrapin, HI18-R5-6, to IAPP compared to HI18 can also be attributed to mutations both directly and indirectly allowing or enhancing hydrogen bonds as well as improving nonpolar interactions to IAPP within the simulations. In subunit 1 (Table IV-1, second column), the I34W and S41N mutations directly enhance hydrogen bonds to IAPP, and the L45V mutation indirectly enhances hydrogen bonds to IAPP by decreasing congestion within the complex structure (Figure IV-6). Specifically, the I34W allows for the formation of a hydrogen bond to N14 of IAPP and enhances nonpolar interactions to V17, N21, and F23 of IAPP (Figure IV-6). Additionally, the S41N mutation enhances the hydrogen bond to N21 of IAPP (Figure IV-6). The L45V mutation supports the I34W mutation by decreasing the congestion that may otherwise be present due to the introduction of the I34W mutation; thereby the L45V mutation indirectly enhances the hydrogen bond between the I34W mutation and N21 of IAPP (Figure IV-6).

In subunit 2, (Table IV-1, third column), the S11Q and A42D mutations directly allow for hydrogen bonds or salt-bridges to IAPP, the A12Q mutation enhances nonpolar interactions to IAPP, and the S39K mutations indirectly enhances HI18-R5-3's affinity to IAPP by stabilizing the unbound state of the designed β-wrapin (Figure IV-6). The S11Q mutation enhances the hydrogen bond and nonpolar interaction to T30 of IAPP (Figure IV-6). This interaction, in addition to the A12Q mutation, also enhances the nonpolar interaction between the N-terminus of subunit 2 and the C-terminus of IAPP (Figure IV-6). Additionally, the A42D mutation allows for the formation of a salt-bridge with K1 of IAPP (Figure IV-6). The S39K mutation supports the A42D mutation and stabilizes the isolated

designed β-wrapin by forming a salt-bridge with A42D (not shown); this is crucial as previous, preliminary studies indicated that a lone A42D mutation would destabilize the unbound β-wrapin. The presence of the S39K mutation in conjunction with the A42D mutation proved beneficial in the simulations as the average RMSF of the isolated HI18-R5-3 is not elevated compared to the isolated HI18 (Table IV-1, fifth column).

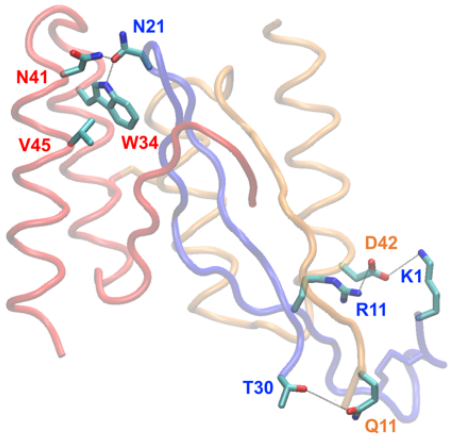

Figure IV-6. Molecular graphics image showing the improved interactions of HI18-R5-6 in complex with IAPP. IAPP is shown in blue tube representation, subunit 1 of HI18-R5-6 is shown in red tube representation, and subunit 2 of HI18-R5-6 is shown in orange tube representation. Blue labels indicate residues of IAPP. Red and orange labels indicate residues of subunits 1 and 2 of HI18-R5-6, respectively.

Design of multi-targeted β-wrapin variants binding Aβ, α-syn, and IAPP with enhanced affinities

The design program was applied in several rounds to design β-wrapin variants with multi-targeted, enhanced affinity to Aβ, α-syn, and IAPP ((2) of Figure IV-2). AS10 was used as a basis as it has the highest affinity to all three amyloidogenic proteins simultaneously[73] (Figure IV-1). MD simulations combined with association free energy calculations and residue-pairwise interaction free energy calculations were used to evaluate the affinity or enthalpic favorability of the designed β-wrapin variants passing the initial screening by the design program ((3) of Figure IV-2). MD simulations combined with RMSF calculations were used to evaluate the stability or entropic favorability of the designed β-wrapin variants ((3) of Figure IV-2).

Based on the design program, MD simulations and free energy calculations of the top-ranked designed β-wrapins in complex with the three amyloidogenic proteins independently, and MD simulations and RMSF calculations of the unbound designed β-wrapins, two promising designed β-wrapins, named AS10-4 and AS10-2 (Table IV-2), were predicted to have enhanced multi-targeted affinity to Aβ, α-syn, and IAPP compared to AS10. Both designed β-wrapins originated from the asymmetric design round in which AS10 subunit 1 residue positions A12, G13, I16, and L45 as well as subunit 2 residue positions I16 and P20 were selected as positions amenable for mutation. The particular residue positions were selected aiming to target corresponding polar residues of Aβ, α-syn, and IAPP,[74, 75] according to the biophysical insights derived in ref. [74, 75]. The targeting of corresponding polar residues of Aβ, α-syn, and IAPP[74, 75] aimed to significantly

improve polar association free energies as previous studies on β-wrapin : amyloidogenic protein complexes, described in the previous sections, suggest that a low polar association free energy in addition to a fairly low non-polar association free energy is characteristic of β-wrapins with high affinity.[74, 75]

The mutations (introduced to the sequence of AS10), energetic favorability (based on ΔΔG compared to AS10), and stability (based on average RMSF) of the two designed β-wrapins are summarized in Table IV-2. In the following text, the predicted improved interactions of AS10-4 and AS10-2 compared to AS10, based on interaction free energy calculations and visual inspection of the six replicate simulation trajectories, are described.

Table IV-2. Summary of the most promising designed β-wrapin variants for multi-targeted Aβ, α-syn, and IAPP binding. ΔΔG values are calculated as the association free energy of the designed β-wrapin variant minus the association free energy of AS10 in complex with the same amyloidogenic protein (Aβ, α-syn, or IAPP). All statistics are averaged over six individual simulation runs.

β-wrapin Variant	**Subunit 1 Mutations to AS10**	**Subunit 2 Mutations to AS10**	**ΔΔG Aβ (kcal·mol^{-1})**	**ΔΔG α-syn (kcal·mol^{-1})**	**ΔΔG IAPP (kcal·mol^{-1})**	**RMSF (Å)**
AS10 (wild type)			0.0 ± 4.2	0.0 ± 2.9	0.0 ± 5.0	2.8 ± 0.1
AS10-2	G13Q, G14H, L45N		-5.4 ± 5.0	-4.33 ± 5.0	-9.1 ± 5.0	2.7 ± 0.6
AS10-4	G13R, G14H, I16T, L45T	I16N, P20N	-15.3 ± 5.0	-8.8 ± 5.0	-17.7 ± 5.0	2.1 ± 0.2

Improved interactions of AS10-4: The predicted enhanced multi-targeted affinity of designed β-wrapin, AS10-4, to Aβ, α-syn, and IAPP compared to AS10 can be attributed to mutations directly allowing or enhancing hydrogen bonds or salt-bridges to Aβ, α-syn, and/or IAPP within the simulations (Figure IV-7). In subunit 1 (Table IV-2, second column), the G14H mutation allows for hydrogen bonds to corresponding polar residues of all three amyloidogenic proteins, the G13R and I16T mutations allow for hydrogen bonds to corresponding polar residues of Aβ and α-syn, and the L45T mutation allows for hydrogen bonds to corresponding polar residues of α-syn and IAPP (Figure IV-7). Specifically, the G14H mutation allows for hydrogen bonds to Aβ D23, α-syn S42, and IAPP H18 (Figure IV-7). The G13R mutation allows for the formation of salt-bridges to Aβ D23 and α-syn E46 (Figure IV-7). The I16T mutation allows for the formation of hydrogen bonds to Aβ E22 and α-syn S42 (Figure IV-7). Additionally, the L45T mutation allows for the formation of hydrogen bonds to α-syn Y39 and IAPP N21 (Figure IV-7).

In subunit 2, (Table IV-2, third column), the I16N and P20N mutations directly allow for the formation of hydrogen bonds or enhance polar interactions to Aβ, α-syn, and/or IAPP within the simulations (Figure IV-7). The I16N mutation allows for the formation of a hydrogen bond to only Aβ M35 (Figure IV-7). The I16N mutation also indirectly enhances the binding of AS10-4 synergistically with the I16T mutation of subunit 1 by forming intramolecular hydrogen bonds with each other (Figure IV-7). Finally, the P20N mutation allows for the formation of hydrogen bonds with the backbone atoms of Aβ I31 as well as with a-syn H50 and IAPP N22 (Figure IV-7).

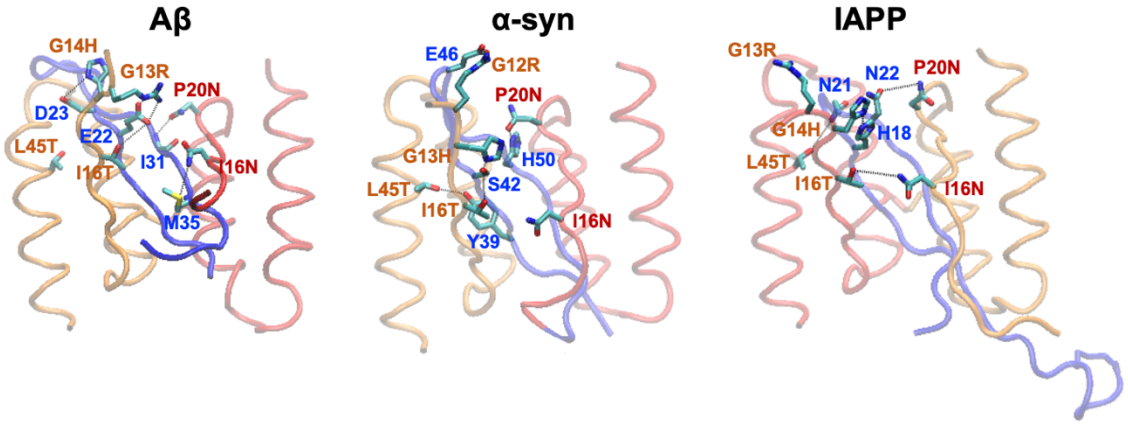

Figure IV-7. Molecular graphics image showing the improved interactions of AS10-4 in complex with Aβ, α-syn, and IAPP. Aβ, α-syn, and IAPP are shown in blue tube representation, subunit 1 of AS10-4 is shown in orange tube representation, and subunit 2 of AS10-4 is shown in red tube representation. Blue labels indicate residues of Aβ, α-syn, or IAPP. Orange and red labels indicate residues of subunits 1 and 2 of AS10-4, respectively.

Improved interactions of AS10-2: The predicted enhanced multi-targeted affinity of designed β-wrapin, AS10-2, to Aβ, α-syn, and IAPP compared to AS10 can be attributed to mutations directly allowing or enhancing hydrogen bonds to Aβ, α-syn, and/or IAPP within the simulations (Figure IV-8). In AS10-2, mutations were only introduced to subunit 1 of the β-wrapin (Table IV-2, second column). The L45N mutation allows for hydrogen bonds to corresponding polar residues of all three amyloidogenic proteins, the I14H mutation allows for hydrogen bonds to corresponding polar residues of Aβ and IAPP, and the G13H mutation allows for a hydrogen bond to a polar residue of α-syn (Figure IV-8). Specifically, the L45N mutation allows for hydrogen bonds to Aβ D23, α-syn K43, and IAPP S19 (Figure IV-8). The G14H mutation allows for the formation of

salt-bridges to Aβ E22 and IAPP H18 (Figure IV-8). Additionally, the G13Q mutation allows for the formation of a hydrogen bond to α-syn K45 (Figure IV-8).

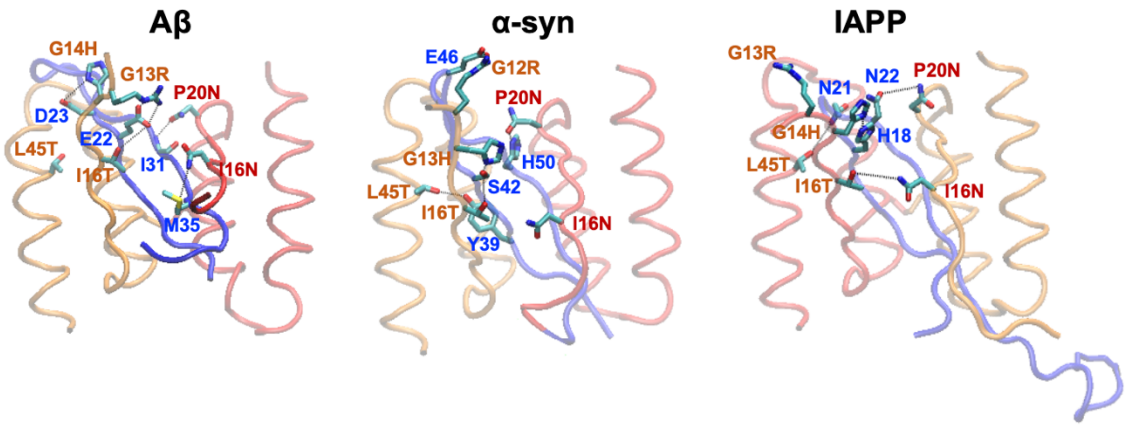

Figure IV-8. Molecular graphics image showing the improved interactions of AS10-2 in complex with Aβ, α-syn, and IAPP. Aβ, α-syn, and IAPP are shown in blue tube representation, subunit 1 of AS10-2 is shown in orange tube representation, and subunit 2 of AS10-2 is shown in red tube representation. Blue labels indicate residues of Aβ, α-syn, or IAPP. Orange labels indicate residues of subunit 1 of AS10-2.

Design of multi-targeted β-wrapin variants binding Aβ and tau with enhanced affinities

The design program was additionally applied to screen and generate dual-Aβ/tau binding β-wrapin variants ((2) of Figure IV-2). As $ZA\beta_3$ has no experimentally detectable binding to tau and TP4 has no experimentally detectable binding to Aβ,[68] an attempt to design β-wrapins which have at least minimal dual-Aβ/tau binding was attempted. Such a β-wrapin with some detectable binding to both amyloidogenic proteins could potentially serve as a steppingstone for the future design of dual-Aβ/tau binding β-wrapin-based therapeutics.

Taking a conservative approach, the design program was asymmetrically applied to introduce mutations to 16 β-wrapin residue positions (8 per β-wrapin subunit), which constituted the only differences between $ZA\beta_3$ and TP4 ($ZA\beta_3$ → TP4: I16M, V17A, Y18S, L19G, F30L, I31V, L45M, and N52S). The 16 residue positions were allowed to switch between $ZA\beta_3$ and TP4 residues ((2) of Figure IV-2). MD simulations combined with association free energy calculations and residue-pairwise interaction free energy calculations were used to evaluate the affinity or enthalpic favorability of the designed β-wrapin variants passing the initial screening by the design program ((3) of Figure IV-2). MD simulations combined with RMSF calculations were used to evaluate the stability or entropic favorability of the designed β-wrapin variants ((3) of Figure IV-2).

Based on the design program, MD simulations and free energy calculations of the top-ranked designed β-wrapins in complex with the two amyloidogenic proteins independently, and MD simulations and RMSF calculations of the unbound designed β-

wrapins, two promising designed β-wrapins, DES16 and DES20 (Table IV-3) were predicted to have detectable and multi-targeted affinity to Aβ and tau. The mutations (to the sequence of $ZA\beta_3$), energetic favorability (ΔΔG compared to $ZA\beta_3$: Aβ and TP4 : tau), and stability (average RMSF) of the designed β-wrapins are summarized in Table IV-3.

Table IV-3. Summary of the most promising designs β-wrapin variants for multi-targeted Aβ and tau binding. ΔΔG values are calculated as the association free energy of the design minus the association free energy of $ZA\beta_3$ in complex with Aβ or TP4 in complex with tau. All statistics are averaged over six individual simulation runs.

β-wrapin Variant	**Subunit 1 Mutations to AS10**	**Subunit 2 Mutations to AS10**	**ΔΔG Aβ : $ZA\beta_3$ (kcal·mol^{-1})**	**ΔΔG tau : TP4 (kcal·mol^{-1})**	**RMSF (Å)**
$ZA\beta_3$ (wild type)			-0.0 ± 3.6	18.1 ± 1.6	2.5 ± 0.1
TP4 (wild type)			7.4 ± 3.8	0.0 ± 1.8	2.3 ± 0.1
DES16	F30L, I31V, N52S	L19G, F30L, L45M	-4.9 ± 3.3	-3.9 ± 1.6	2.4 ± 0.2
DES20	F30L, I31V	F30L, L45M	1.7 ± 2.0	-5.7 ± 6.0	2.3 ± 0.4

Rational Design of β-wrapins Incorporating Non-canonical Amino Acids for Bifunctional Fibril Deformation and Monomer Sequestration Properties

In the design of β-wrapins described in the previous sections, the computational design was focused on introducing mutations to the core of the β-wrapins, which sequester amyloidogenic protein monomers and induce a β-hairpin fold. In addition to the computational design of β-wrapin variants with enhanced multi-targeted binding to Aβ, tau, α-syn, and/or IAPP, an alternative direction for inhibiting the aggregation of amyloidogenic proteins was explored. In this alternative direction, the flexible, currently "non-functional" N-terminus of β-wrapins was hypothesized to be modifiable for amplified anti-amyloid properties. Specifically, aromatic canonical and non-canonical amino acids were considered possible "decorations" that could be incorporated into the N-termini to mimic compounds with anti-amyloid properties. The rational design of N-termini was focused on a polyphenolic compound, curcumin, which has been suggested to have protective properties against AD,[281-283] PD,[284-286] and T2D.[287-289] Curcumin has been shown to modulate the aggregation of Aβ,[283, 290] tau,[291, 292] α-syn,[286, 293] and IAPP.[294, 295] Furthermore, curcumin has also been shown to reduce amyloid *in vivo*[283] and disintegrate preformed tau filaments *in vitro*.[292] Due to its properties,[281-292, 294, 295] curcumin appears to be a promising steppingstone for the design of novel agents targeting AD, PD, and T2D. Thus, a β-wrapin with the ability to mimic curcumin's amyloid inhibition mechanisms was designed and hypothesized to have amplified anti-amyloid properties, comprising the anti-amyloid properties of both the β-wrapin core and curcumin.

In Dr. Tamamis' lab, curcumin, and a set of curcumin derivatives were investigated in complex with a hexamer peptide Aβ fibril model through nearly exhaustive docking, followed by multi-ns MD simulations and structural and energetic analysis to provide insights into their Aβ fibril binding and Aβ aggregation inhibitory properties.[296] In a portion of the simulations, the compounds partly dissociated the outermost peptide of the $A\beta_{1-42}$ fibril by disrupting β-sheets within the residue domain $_{12}$VHHQKLVFF$_{20}$ through specific interactions leading to partial dissociation.[296] These simulations suggested potential inhibition mechanisms of $A\beta_{1-42}$ aggregation by the compounds: the partially dissociated $_{16}$KLVFF$_{20}$ domain of outermost peptide could either remain unstructured or wrap around to form intramolecular interactions with the same peptide's $_{29}$GAIIG$_{33}$ domain;[296] the compounds could additionally act as a patch against the external edge of the second outermost peptide's $_{16}$KLVFF$_{20}$ domain.[296] Thereby, individually or concurrently, these could prohibit fibril elongation.[296]

Inspired and guided by Dr. Tamamis' lab's studies on curcumin,[296] an endeavor to amplify the anti-amyloid properties of β-wrapins was undertaken based on the hypothesis that the N-termini of β-wrapins, which do not have a particular role in binding according to NMR studies,[69-71] could be modified by introducing aromatic canonical and non-canonical amino acids that structurally and physicochemically mimic the chemical groups of curcumin. Such modifications could confer the ability of curcumin to dissociate external peptides of Aβ fibrils[296] to the designed β-wrapins and potentially protect the β-wrapins from protein degradation. Thus, such modifications to the β-wrapins' N-termini

were hypothesized to constitute an alternative means to improve their binding and inhibitory properties.

Molecular docking of the N-terminus of β-wrapin AS10 to a modeled Aβ hexamer fibril using ZDOCK[297] showed that the first five residues could fit and bind to the same region that curcumin binds according to simulations[296] and experiments[298] (Figure IV-9). Specifically, AS10 residues 1 and 5 were located in similar positions as the functional "dissociator" and "anchor" groups, respectively, of curcumin as defined in previous studies[296] (Figure IV-9). Aiming to rationally design the β-wrapin AS10 N-termini to mimic curcumin, combinations of aromatic canonical and non-canonical amino acids were introduced to AS10 residues 1 and 5. Upon introduction of aromatic canonical and non-canonical amino acids, resulting in 26 investigated AS10 N-termini (25 rationally designed, modified N-termini and the unmodified AS10 N-terminus), the rationally design N-termini were subsequently investigated in complex with the Aβ hexamer fibril using 100 ns MD simulations and structural analysis. Among the AS10 N-termini investigated in complex with the Aβ hexamer fibril, triplicate 100 ns MD simulations of two modified N-termini, FDNK-(MOT5)-NKEMA and FDNK-(NAO2)-NKEMA, showed that the two N-termini could mimic the Aβ fibril elongation inhibition mechanism of curcumin by partly dissociating the outermost Aβ peptide of the fibril (Figure IV-10).

For both modified N-termini, the combination of a mutating the first residue position to Phe and the fifth position to particular aromatic non-canonical amino acids enhance specific interaction leading to the partial dissocation of the Aβ fibril. The canonical Val1Phe mutation enhanced nonpolar interactions to outermost Aβ peptide

Val12 and Leu17 as well as π–π interactions with outermost Aβ peptide His14 and Phe19 compared to the unmodified AS10 N-terminus; this is analogous to the interactions of cucumin's "dissociator" group with the outermost Aβ peptide that result in partial dissociation of the bound Aβ fibril, which is suggested to inhibit amyloid elongation.[296] Additionally, for both modified N-termini, the non-canonical mutation of Phe5 to either MOT5 or NAO2 enhanced their nonpolar interactions to interior Aβ peptides Ile32, Gly33, and Leu34 and allowed for the formation of hydrogen bonds with the backbone atoms of interior Aβ peptides Gly33 or Leu34; this is analogous to the interactions of cucumin's "anchor" group with the Aβ peptides in the interior of the fibril that result in partial dissociation of the bound Aβ fibril, which is also suggested to inhibit amyloid elongation.[296] Additionally, association free energy calculations suggested that the partial dissociation of the outermost peptide of the Aβ fibril by the two modified AS10 N-termini could be an outcome of either high affinity interactions or a cause leading to high affinity interactions between the N-termini and the Aβ fibril, which could partly serve as a compensation for the energy loss between the outermost peptide and the rest of the fibril due to partial dissociation, in line with ref. [296]. Importantly, even in the context of the full β-wrapin, both modified N-termini partially dissociate the outermost peptide of the Aβ fibril without inhibiting the ability of the β-wrapin's core to sequester the amyloidogenic protein monomer.

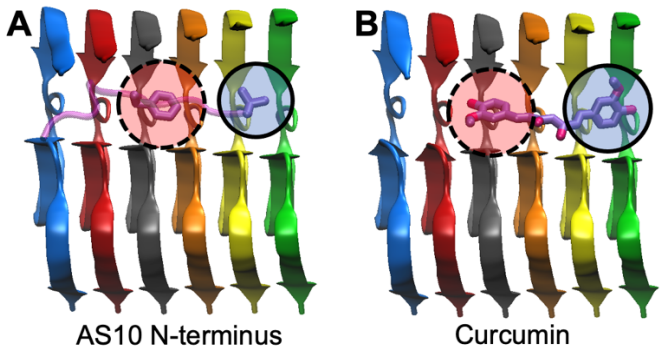

Figure IV-9. Molecular graphics images of (A) unmodified AS10 N-terminus and (B) curcumin in complex with the Aβ fibril. The structure of (B) curcumin in complex with the Aβ fibril corresponds to a docked structure that induced the partial dissociation of the outermost peptide of the Aβ fibril derived from ref. [296]. The red shaded circle indicates the overlap of AS10 residue 5 with the "anchor" group of curcumin, and the blue shaded circle indicates the overlap of AS10 residue 1 with the "dissociator" group of curcumin.

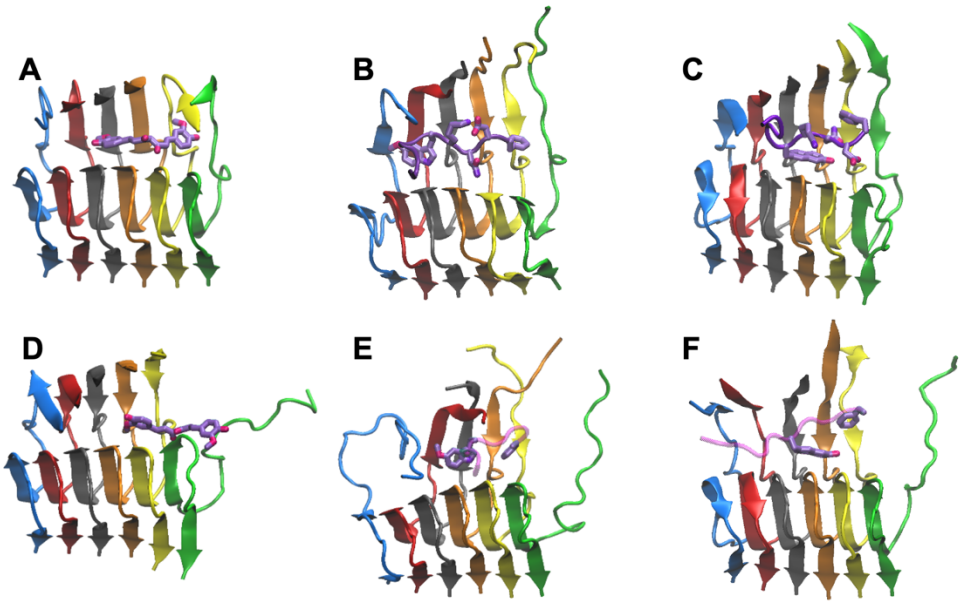

Figure IV-10. Molecular graphics images of a (A,D) curcumin in complex with the Aβ fibril, (B,E) modified N-terminus, FDNK-(MOT5)-NKEMA, binding to the Aβ fibril, and (C,F) modified N-terminus, FDNK-(NAO2)-NKEMA, binding to the Aβ fibril (A,B,C) prior to the initiation of partial dissociation of the Aβ fibril and (D,E,F) at 100 ns. The structures of curcumin in complex with the Aβ fibril are derived from ref. [296].

Concluding Remarks for the Study and Design of β-wrapins Binding to Amyloidogenic Proteins

The self-assembly of amyloidogenic proteins Aβ, tau, α-syn, and IAPP into fibrils is associated with amyloid diseases AD, PD, and T2D.[232-239] β-wrapins bind to and sequester monomers of amyloidogenic proteins, thereby inhibiting their aggregation. [21, 68-73] Existing β-wrapins have traditionally been engineered with single-targeted, micro-molar affinities to the amyloidogenic protein monomers through phage-display libraries, with mutable β-wrapin positions selected based on intuition and insights from NMR studies.[21, 68-73] From phage-display libraries aiming to discover β-wrapins for α-syn, AS10 was discovered to have multi-targeted Aβ, α-syn, and IAPP micro-molar affinity.[73] Despite the successes of experimental methods in discovering β-wrapins,[21, 68-73] the design of high affinity β-wrapins with dual- Aβ/tau and triple- Aβ/α-syn/ IAPP binding properties remains challenging.

In this chapter, MD simulations, free energy calculations, and structural analysis have been carefully tuned for the study of β-wrapins in complex with amyloidogenic proteins. The studies described in this section show that these methods can be applied across different amyloidogenic proteins to model their interactions with β-wrapins.[74, 75] From these MD simulation studies,[74, 75] specific interactions and energetic driving forces leading to β-wrapin single- and multi-targeted binding properties were uncovered. Additionally, the novel protein design model developed for this chapter have produced promising β-wrapin designs that according to MD simulations and free energy calculations have predicted enhanced affinity for their respective amyloidogenic protein(s);

nevertheless, experiments are needed to validate the computational predictions. Finally, the promising results of the modification of β-wrapin N-termini with aromatic canonical and non-canonical amino acids inspired by curcumin's binding and inhibitory function to Aβ could provide impetus for future studies of modified N-termini in complex with additional amyloid fibrils.

CHAPTER V

LIGANDS BINDING TO CLAYS

Simulation Studies of Toxic Compounds Binding to Montmorillonite Clays

Introduction

Since ancient times, peoples of several cultures have identified and used clays as medicine for a variety of ailments including infection, digestive disease, and poisoning.[299-301] In modern times, clays find use in medicine and have proved their usefulness in the treatment of skin conditions,[302-308] bacterial infections,[309-312] digestive disease,[309, 311, 313] and poisoning[314] among others. The physical properties of clay minerals that make them medicinal materials (large surface area, chemical and mechanical stability, high adsorptive properties, high cation exchange capacity, etc.[315, 316]) also make then a natural scavenger of pollutants and, thereby, strong candidates as adsorbent materials for the removal of various toxic chemicals.[24, 25] On this front, montmorillonite clays have been reported to tightly bind various toxic chemicals within their interlayers, including heavy metals,[317, 318] pesticides,[77, 319-321] and mycotoxins.[322-324] Furthermore, a series of intervention studies showed that montmorillonite clay is safe for short-term human[325-327] and animal[328-331] consumption. Due to its ability to bind toxic chemicals, montmorillonite clays are promising candidates to serve as adsorbents to mitigate exposure to toxic chemicals in the wake of environmental emergencies, during which the risk of toxic chemical exposure cab be amplified for first responders and afflicted people or animals.[24, 25]

Experimental methods have provided valuable insights into the structure of montmorillonite clays and their interactions with chemicals. X-ray diffraction reveal the d-spacing of montmorillonite clay layers and give insights into the clay layers in aqueous solutions;[332] Infrared spectroscopy monitor conformational changes of surfactants intercalated into montmorillonite interlayers;[333] Differential thermal analysis give thermodynamic insights into the dehydration, dihydroxylation, amorphization, and rehydration of montmorillonite;[334] Thermogravimetric analysis of montmorillonite composites can be used to evaluate its thermal stability and the adsorption of water;[335, 336] Electron microscopy provide images of montmorillonite clay, providing further insights into the layer thickness, lattice structure, degree of crystallinity, and turbostratic disorder of montmorillonite clay.[337, 338] Several experimental methods also exist to evaluate the ability of montmorillonite clays to adsorb chemicals. Isothermal adsorption studies of montmorillonite in chemical gradient solutions estimate the maximum capacity and thermodynamic properties of binding of the clay for a given chemical.[339] The level of accuracy obtained from adsorption processes is greatly dependent on the successful modelling and interpretation of adsorption isotherms.[340] Generally, Langmuir and Freundlich are the most examined models in fitting the experimental adsorption data, and the adsorption kinetics is predominantly based on the pseudo-second-order model.[341] The detoxification efficacy of clay-based adsorbents and their ability to protect living organisms from chemical exposures have also been investigated through *in vivo* studies.[318, 322, 324, 328-330, 342]

Computational methods have also been valuable tools to model the interaction between clays and compounds. Computationally, an atomistic-level examination of the dynamics and the delineation of molecular mechanisms of compound adsorption onto clays is possible. Quantum mechanics (QM) simulations explicitly include electron interactions and are able to model reaction mechanisms between clay and organic compounds, brine, and water.[343-346] Using QM methods, the extent of interaction between compounds and clay surfaces, potential reactivity of a system, as well as energetic favorability of potential reaction mechanisms can be evaluated and compared across like systems.[343-346] Additionally, structures resulting from QM simulations show how compounds may orient themselves with respect to clay surfaces upon adsorption.[344-346] Despite the highly detailed insights that can be gained from QM methods, the complexity of composition and structure of most clays and the sheer number of potential interactions between clays and compounds make QM methods less attractive due to their restricted time- and length-scales.[347]

Atomistic MD dramatically increase both time- and length-scales that can be simulated compared to QM by treating atoms as spheres intramolecularly bonded to one another via a set of springs. Because of this, classical MD methods is an attractive and popular approach to model clay systems.[37, 346-353] MD simulations are able to sample and probe the adsorption of compounds onto different regions of clays (e.g., interlayer basal surfaces or edge surfaces) to provide fundamental, atomistic insights into clay : compound interactions. Structures of compounds bound to clay surfaces derived from MD simulations and comparisons to experimental data have provided valuable insights into

the interactions and orientation of the compounds with the clay.[346, 347, 349, 350, 353] Structural and energetic data extracted from MD simulations have also been used to predict adsorption free energies of compounds for clays, with the methods used to predict adsorption free energies tailored to single types of compounds and clays.[37, 348, 351, 352] The methods used in these studies vary widely;[37, 346-353] The duration of MD simulations used range from very short (50 ps to 20 ns)[346, 348-352] to long (100 ns or longer),[37, 347, 353] with a compound's ability to bind to edge surfaces of the clay either considered - allowing compounds to enter and escape clay interlayers,[37, 348, 351] or not considered - disallowing compounds from leaving the clay interlayers.[346, 348, 350] Overall, previous MD simulation[37, 346-353] and experimental[319-324] studies have demonstrated that many toxic compounds can bind tightly to clays.

In this chapter, the systematic generation of MD simulation files for simulating and investigating the binding of compounds onto montmorillonite clays is described. The systematic generation of simulation systems contributed to the comparison of diverse compounds' propensity to bind to clay,[76-78] and ultimately enabled the development of a minimalistic model that can predict the adsorption free energy of diverse toxic compounds for CM.[79] Upon completion of the simulations, the simulation snapshots can be further examined through structural and energetic analysis to gain insights into the molecular mechanisms of specific compounds binding to unamended (parent) and nutrient-amended montmorillonite clays. The results of these studies had reasonably high agreement with experiments.[76-79] Comparisons between the binding propensities of different toxic compounds to the clay surfaces were used to delineate toxic compounds that can be

effectively adsorbed onto the clay from those that cannot.[76-78] Additionally, key modes of binding were also identified for the investigated compounds binding to the clay.[76-79] Finally, a minimalistic model was developed to use data extracted from MD simulations as input to output predicted adsorption free energies for toxic compounds to unamended clay.[79]

Methodology for Simulating and Investigating the Binding of Compounds onto Montmorillonite Clays

To simulate and investigate the mechanisms of toxic compounds binding onto montmorillonite clays, MD simulations followed by structural and energetic analysis are performed. Within the simulations, clays are simulated in the presence of toxic compounds to sample their interaction. Structural and energetic analysis subsequently use the simulation snapshots as input to examine the snapshots in which the toxic compound is interacting with the clay. This chapter focuses on the simulation of montmorillonite clays in the presence of toxic compounds, as well as nutrients considered as amending molecules; specifically, the simulation setup was standardized to allow the comparison of diverse compounds binding to montmorillonite clays. The programs producing standardized simulation files can accommodate different montmorillonite clays (Ca^{2+} montmorillonite, Na^{2+} montmorillonite) at different pH conditions (acidic pH 2 and neutral pH 7), different amending molecules on the clay surfaces (L-carnitine, choline, among others), different solvents (water, methanol, acetonitrile, mixtures, among others), different toxic compounds of interest, different counterions within the system, and different sizes of the simulation system (one-layer of clay or two-layers of clay). Figure V-1 shows the options that can be selected for the standardized MD simulations (boxed in red) and the simulation file outputs (boxed in blue). Based on a user defined adsorbent, amending molecule, solvent, counter ion, compound of interest, and simulation system size, the simulation files are automatically generated (Figure V-1). The generalizability and automatic generation of input files is a key step to standardize the simulations and

provide the ability to simulate and adsorbent : toxic compound/amending molecule system under varying conditions. Based on user selections (adsorbent, amending molecule, solvent, counter ion, compound of interest, and simulation system size) simulation files are generated, and appropriate structures and force field topology and parameters are extracted from a repository to build the simulation system (Figure V-1). The automated generation of simulation input files is inspired by and based on CHARMM-GUI.[274] All generated simulation files are formatted for use with CHARMM.[197] After the execution of the simulation files, the coordinate of the equilibrated adsorbent : compound system and the trajectory of the compounds binding to the adsorbent are produced for subsequent analysis. The molecular modeling and MD simulations are described in detail in refs. [76-79].

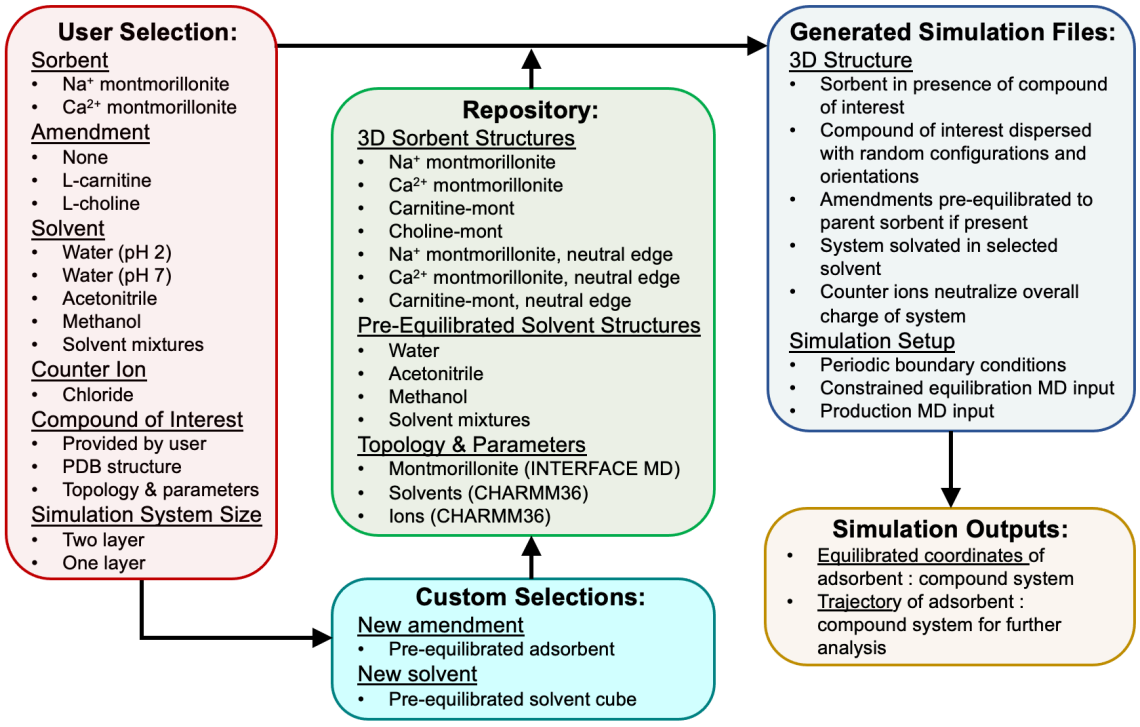

Figure V-1. Schematic view of automated simulation setup for simulating the binding of toxic compounds onto montmorillonite clays.

Stage 1: Selection of Simulation Size and Adsorbent

The simulation input files require an initial structure of the adsorbent, with appropriate amending molecules, in the presence of the compound of interest in addition to appropriate molecular mechanics force field topologies and parameters. A repository provides initial coordinates as well as topologies and parameters for parent, or unamended, adsorbents and amended adsorbents. From this repository, a user can additionally select a two-layered clay (comprising two parallel slabs of clay layers within the simulation system)[76, 78] or a one-layered clay (comprising just one slab of clay layer within the simulation system).[79] Currently, the available adsorbents in the repository for two-layer

simulation systems are Na^+ montmorillonite (SM), Ca^{2+} montmorillonite (CM), L-carnitine-amended CM (CM-carnitine), choline-amended CM (CM-choline), SM with neutral edges (pH 3 and below), CM with neutral edges (pH 3 and below), and CM-carnitine with neutral edges and protonated carnitine molecules (pH 3 and below). For one-layer simulation systems, SM, CM, SM with neutral edges (pH 3 and below), and CM with neutral edges (pH 3 and below) are currently available in the repository (Figure V-1, boxed in green).

The structures of the parent montmorillonite clays were built using atomic coordinates from the INTERFACE model database,[354] and the topology and parameters for the clays were extracted from the INTERFACE force field.[354, 355] The INTERFACE force field can operate as an extension of several commonly used harmonic force fields, including CHARMM,[197] thus enabling simulations of toxin binding on organic/biomolecular and inorganic interfaces,[356-361] including montmorillonite.[76-79, 355] For the two-layer simulation system, the structures of the parent montmorillonite clays correspond to two independent clay layers with a d_{001} spacing of 21 Å.[324, 362] For the one-layer simulation system, the structures of the parent montmorillonite clays correspond to one layer. In both the two- and one-layered simulation systems, periodic boundary conditions within the MD simulations replicate the d_{001} spacing of 21 Å[324, 362] such that both simulation systems represent infinite layers of clay.[76-79]

The structures of the amended montmorillonite clays correspond to MD equilibrated structures of the two-layer parent montmorillonite clay in the presence of copies of the amending molecule. Currently, CM-carnitine at pH 7, CM-choline at pH 7,

and CM-carnitine at pH 2 are readily available in the repository (Figure V-1, boxed in green). If the structure of the desired amended clay is not readily available in the repository, simulation files are generated to produce the model in line with the methods used to generate the models of CM-carnitine and CM-choline (Figure V-1, boxed in cyan). The methodology to introduce amending molecules to the parent clay to model the amended aims to simulate experimental conditions; experimentally, the L-carnitine and choline amended clays were formed by mixing the parent clay with the amending molecule (L-carnitine or choline) in solution and subsequently centrifuging and washing the resulting amended clay with distilled water.[76, 78] Thus, the amended montmorillonite clay models are modeled by first simulating the parent clay in the presence of copies of the amending molecule for a short 10 ns equilibration, representing the initial mixing of the parent clay with amending molecules in experiments, and then extracting the final 10 ns simulation snapshot coordinates of the clay and the individual amending molecules bound to the clay.[76, 78] The simulation setup for the equilibration runs of the parent clay and new amending molecules is the same as the setup for investigating parent clays in the presence of a compound of interest (described below), with the exception that the production run is only 10 ns in duration. Finally, the 10 ns structure is saved as the initial structure of the amended clays, aiming to represent the washed amended clay in experiments. This equilibrated amended clay is subsequently stored in the repository for future use. For the amended clays currently available in the repository, CM-carnitine (at pH2 and pH 7) and CM-choline (at pH 7), the topology and parameters for L-carnitine and

choline were generated using CGenFF[193-196] and molecular structures extracted from the ZINC database.[363]

Stage 2: Solvating the Simulation System and Introducing Compounds of Interest

The selected adsorbent is subsequently solvated in a solvent box with the compound of interest distributed throughout the solvent box in random configurations and orientations. Unless explicitly specified by the user,[77] the dimensions of the solvent box are defined such that, with the periodic boundary conditions applied in the MD simulations, the modeled adsorbent would have infinite layers with an d_{001} spacing of 21 Å between each adsorbent layer.[76, 78, 79] For the two-layer simulation system, this corresponds to a solvent box size of 90 × 90 × 54 $Å^3$.[76, 78] For the one-layer simulation system, this corresponds to a solvent box size of 90 × 90 × 21 $Å^3$.[79] The solvent box is built by replicating pre-equilibrated solvent cubes, from the repository, until the full solvent box is constructed. The user may select a readily available solvent from the repository or build a custom pure or binary solvent. Currently, the repository contains pre-equilibrated solvent cubes for water (extracted from CHARMM-GUI[274]), acetonitrile, methanol, and solvent mixtures (50:50 acetonitrile:water, 60:40 acetonitrile:water, 65:35 acetonitrile:water, 70:30 acetonitrile:water, 90:10 acetonitrile:water, and 90:10 methanol:water). The topologies and parameters for all solvents currently available in the repository correspond to the topology and parameters in the CHARMM36 force field[42, 198] (Figure V-1, boxed in green). In the case that the user selects a custom solvent, and the pre-equilibrated cubes of the solvent under investigation is not readily available in the

repository, simulation files to build the pre-equilibrated solvent cubes can be generated (Figure V-1, boxed in cyan). For pure solvents, the user provides the molecular structure, density, and molecular weight of the solvent under investigation as well as its topology and parameters. Using this information, solvent molecules are placed in a 20x20x20 $Å^3$ cube. The number of solvent molecules placed within the cube is determined based on the solvent's density and the solvent molecule's molecular weight. Short 5 ns MD simulations are subsequently performed to pre-equilibrate the solvent cube. The final 5 ns simulation snapshot is subsequently saved and put in the solvent repository. For solvent mixtures, pre-equilibrated solvent cubes of the pure components of the mixture are first generated as previously described. Then, according to the composition of the solvent mixture, a pure portion of one component of the solvent mixture is combined with other pure portion of the solvent mixture. For example, in the case of the 90:10 by volume acetonitrile:water solvent mixture, pure acetonitrile and pure water solvent cubes were first produced. Using the pre-equilibrated pure acetonitrile and pure water solvent cubes, 10% of the water molecules by volume within the water cube were combined with the 90% of the pure acetonitrile solvent cube. Subsequently, short 5 ns MD simulations are performed to pre-equilibrate the solvent mixture cube.

For the compound of interest, the user provides an initial molecular structure of the compound as well as the compound's topology and parameters. The initial molecular structure as well as topology and parameters are used to generate random configurations of the compound of interest. Care should be taken to provide the correct protonation state of the compound's initial molecular structure for the environment of the simulation

system. It is worth noting that the pH value at the surfaces of clay minerals can be 2 to 4 units below the pH value of the solution.[364, 365] Within the simulations, the pH of the system under investigation is accounted for by simulating the clays and toxic compounds/amending molecules with the corresponding protonation states according to the pK_as of their protonation sites. For example, in the study of CM and CM-carnitine in the presence of PFOA at pH 2, the model of CM used corresponded to CM with neutral edges, the model of carnitine used corresponded to carnitine with its carboxyl group protonated, and the model of PFOA corresponded to its carboxyl group both protonated and unprotonated (in separate simulation systems).[78] In this study, the use of both the protonated and unprotonated forms of PFOAs were used a wide range of pK_a values have been reported for PFOA (ranging from -0.5 to 3.8).[78, 366-368]

The random configurations of the compound of interest are generated from short 1 ns simulations of a single molecule at infinite dilution using the GBSW implicit solvent model.[133] For the two-layer simulation system, 16 molecules of the compound of interest are introduced to the simulation system by default.[76, 78] For the one-layer simulation system, 3 copies of the compound of interest are introduced to the simulation system by default.[79] The number copies of the compound introduced can be increased or decreased depending on the desired simulated density.[77] It is worth noting that while the simulation files generated are specific for the study of only one type of toxic compound/amending molecule, the simulation files generated can be used as a basis to investigate mixtures of compounds. For example, in a system containing 8 copies of hypothetical compound 1 and 8 copies of hypothetical compound 2, the simulation setup can be independently

performed for both hypothetical compounds and the coordinates of alternating odd hypothetical compound 1 copies can be saved (for example, copies 1, 3, 5, 7, 9, 11, 13, and 15) and alternating even hypothetical compound 2 copies can be saved (for example, copies 2, 4, 6, 8, 10, 12, 14, and 16). In this way both hypothetical compounds would be evenly dispersed around the adsorbent under investigation.

Counterions may be added to the solvated system to neutralize the overall charge of the system. The numbers of ions are automatically determined by the total charge of the system. The initial configuration of added ions is determined through short Monte Carlo (MC) simulations (2000 steps).[274] Topologies and parameters for several ions are available in the CHARMM36 force field.[42, 198]

Stage 3: MD Simulations of the Binding of Compounds onto Adsorbents

Prior to the production stage of the MD simulation, the simulation system is initially energetically minimized through 500 steps of steepest gradient descent minimization, 500 steps of Newton-Rapson minimization, and 500 steps of steepest descent minimization followed by a constrained 1 ns MD simulation equilibration stage. During the energy minimizations and 1 ns equilibration stage, the adsorbent layer(s), amending molecules (if present), and compounds of interest are constrained with a 1.0 $kcal \cdot mol^{-1} \cdot Å^{-2}$ harmonic constraint on all heavy atoms. Following energy minimization and equilibration, the simulation system enters the production stage with all constrains on the system are released except, for the montmorillonite clay adsorbents, light 0.1 $kcal \cdot mol^{-1} \cdot Å^{-2}$ harmonic constraint on aluminum atoms of the clay layers. The duration

of the production stage can be lengthened or shortened as desired.[76-79] For example, in studies involving two-layer montmorillonite clay systems in the presence of compounds involved in the production of plastics, bisphenols and phthalates,[76] the production stage duration was 100 ns; in the study involving two-layer montmorillonite clay systems in the presence of herbicides, glyphosate and paraquat,[77] as well as PFAs,[78] the production stage duration was 50 ns; in the study involving a one-layer montmorillonite clay system in the presence of diverse toxic compound in different solvents,[79] the production stage duration was 30 ns. In the simulations, the temperature and pressure are set to 300 K and 1 atm, respectively, hydrogen bond lengths are constrained using the SHAKE algorithm,[278, 279] and periodic boundary conditions are applied. MD simulation snapshots are extracted in 20 ps intervals for subsequent analysis (Figure V-1, boxed in yellow).

Simulation snapshots have been used as input to structural analysis and energy calculation programs. In the structural analysis programs developed by members of Dr. Tamamis' lab,[76-78] simulation snapshots in which the toxic compound is bound to the clay (based on distance cutoffs between atoms of the compound and the clay or amending molecule) are collected. Based on the number of instances the toxic compound is bound to the clay (either through direct interaction with the clay, indirect interaction with the clay through interactions with amending molecules, or simultaneous interactions directly with the clay and with amending molecules) within the simulations, binding propensities of the toxic compound are calculated. Based on which groups of atoms of the toxic compound are most frequently bound to the clay and/or amending molecule in these instances, the key binding modes of the compound are detected.

In the energy calculation programs, simulation snapshots of the toxic compound binding to the clay are used as input for interaction energy calculations. Such energy calculations can provide insights into the driving forces leading to the toxic compounds binding. In the energy calculations, the electrostatic and van der Waals interaction energy of the toxic compounds binding to the clay are calculated. In the cases in which the toxic compound is directly bound to the clay surface (without interacting with any amending molecule in the case of amended clays), the interaction energy of the toxic compound to the clay is calculated.[76, 79] In the cases in which the toxic compound is either indirectly bound to the clay through interactions with amending molecules, or simultaneously bound to both the clay and amending molecules, then the interaction energies are calculated between the bound toxic compound and the amended clay, where only the amending molecule involved in the binding of the toxic compound is included as part of the amended clay.[76] All energy calculations are performed in CHARMM[197] using infinite cutoffs.[76, 79] As is, it is important to note that the interaction energy values show the strength of interaction between the toxic compound and the clay and do not represent the absolute adsorption free energy of the toxic compounds for the clay. To predict the adsorption free energies of diverse toxic compounds binding to clays, these interaction energy values were used as input to a minimalistic model (described in the following section).[79]

Applications of Simulating and Investigating the Binding of Compounds onto Montmorillonite Clays

The methodology for simulating and investigating the binding of compounds onto montmorillonite clays has been used in several studies investigating the mechanisms of diverse toxic compounds binding to montmorillonite clays. The standardization of the simulations has allowed qualitative comparison across structurally and physicochemically distinct compounds as well as across different clay formulations (SM, CM, CM-carnitine, CM-choline).[76-78] Further the standardization of the simulations and energy calculations have also allowed for the development of a minimalistic model that predicts the adsorption free energy of individual compounds for CM.[79] Experiments for these studies were performed by members of Dr. Phillips' lab at Texas A&M University, College Station. Contributors to these studies are listed in the authors lists of refs. [76-79]. Additional details of these studies are provided in refs. [76-79].

Interactions of Na^+ Montmorillonite with Herbicides, Glyphosate and Paraquat

Glyphosate and paraquat are two of the most widely used herbicides.[369, 370] Due to their common usage, they are ubiquitous contaminants in the environment that can affect human health.[371-385] During events of extreme weather such as hurricanes or floods, soils contaminated with glyphosate or paraquat can be mobilized and redistributed to areas of high human and animal contact.[386] In this study, the binding mechanisms of glyphosate and paraquat to SM at acidic (pH 2) and neutral (pH 7) pH conditions were computationally investigated through MD simulations and structural analysis.[77]

According to MD simulations and experiments, both glyphosate and paraquat can be tightly bound by SM.[77] In the simulations, glyphosate was primarily bound to SM through its positive amide group at pH 2 and 7 (51% and 74% of all binding instances, respectively, Figure V-2A,B).[77] Due to protonation at pH 2, glyphosate was also bound to SM through hydrogen bond interactions between its phosphate group and the oxygens of the clay siloxane surface (43% of all binding instances, Figure V-2B).[77] Paraquat was primarily bound to SM with both bipyridinium rings being approximately parallel to the surface at both pH 2 and pH 7 (90% and 93% of all binding instances, respectively, Figure V-2C,D).[77] These results are consistent with previous experimental studies showing that glyphosate and paraquat bind strongly with soils containing clays.[387-389] Paraquat was also observed to bind to the clay with longer residence times per binding event compared to glyphosate suggesting that paraquat has a higher probability to retain a bound conformation compared to glyphosate at pH 2 and 7.[77] The computationally calculated average residence time per binding event of glyphosate bound to clay at pH 2 and pH 7 were 0.59 ± 1.2 ns and 0.10 ± 0.09 ns, respectively.[77] The computationally calculated average residence time per binding event of paraquat bound to clay at pH 2 and pH 7 were 2.84 ± 6.33 ns and 4.06 ± 7.45 ns, respectively.[77] These results are in line with experimental hydra assays suggesting that the bioavailability of paraquat is lower than that of glyphosate when in the presence of montmorillonite clay.[77]

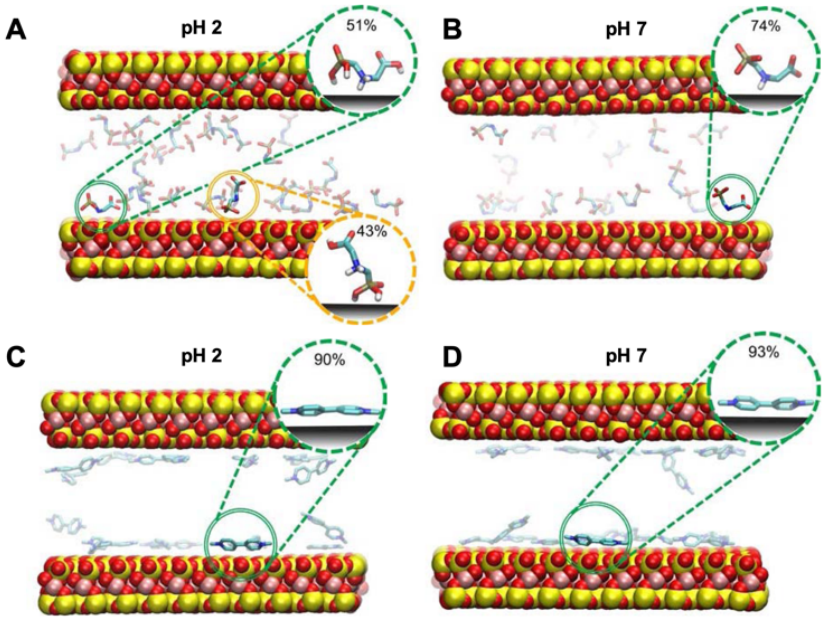

Figure V-2. Molecular graphics images from representative MD simulation snapshots of (A,B) glyphosate and (C,D) paraquat binding to SM at (A,C) pH 2 and (B,D) pH 7. The most prominent binding conformations of (A,B) glyphosate or (C,D) paraquat and their corresponding propensities are encircled in green dotted lines adjacent to the image of the simulation system and green doubled lines within the image of the simulation system. (A) The second most prominent binding conformation of glyphosate at pH 2 is encircled in orange dotted lines adjacent to the image of the simulation system and orange doubled lines within the image of the simulation system. This figure is an adaptation from ref. [77*].

---

* Reprinted with permission from "Montmorillonites Can Tightly Bind Glyphosate and Paraquat Reducing Toxin Exposures and Toxicity" by Meichen Wang, Asuka A. Orr, Shujun He, Chimeddulam Dalaijamts, Weishueh A. Chiu, Phanourios Tamamis, and Timothy D. Phillips, 2019. *ACS Omega*, 4 (18), 17702-17713, Copyright 2019 by American Chemical Society. https://pubs.acs.org/doi/10.1021/acsomega.9b02051. Further permissions related to the material excerpted should be directed to the ACS

Interactions of Bisphenols and Phthalates with Parent and Carnitine-Amended Ca^{2+} Montmorillonite Clays

Bisphenols, bisphenol A (BPA) and bisphenol S (BPS), as well as phthalates, dibutyl phthalate (DBP) and di-2-ethylhexyl phthalate (DEHP), are commonly used in the production of plastics.[390-394] BPA and BPS have been suggested to be xenoestrogens, exhibiting estrogen-mimicking, hormone-like properties, and have been linked to endocrine disorders.[394-401] DBP and DEHP have also been suggested to be endocrine-disruptive chemicals based on both animal[402-405] and human[406-409] studies. Due to the pervasive use of plastics in modern technologies ranging from food storage to medicine,[392, 410] the exposure of vulnerable populations to potentially harmful chemicals that can be leached from plastics, such as bisphenols and phthalates, is a concern.[393, 394, 396, 406] In this study the binding mechanisms of bisphenols and phthalates to both parent (unamended) CM and CM-carnitine in acetonitrile was investigated.[76]

According to MD simulations and binding propensity calculations, the parent CM was predicted to be capable of binding BPA and BPS but not effective in binding DBP or DEHP.[76] Additionally, while the addition of carnitine amendments to CM, resulting in CM-carnitine, improved the binding of the clay to BPA and BPS compared to the parent CM, it did not improve the binding of the clay to DBP or DEHP.[76] Interaction energy calculations indicated the strength of interaction of BPA and BPS to the clay and suggested that electrostatic interactions were the key driving force leading to their binding to the clay.[76] In line with these computational results, experiments comparing the binding of CM to BPA and DBP confirmed that the clay was capable of binding BPA, but not DBP.[76]

Additionally, experiments examining the effect of carnitine-amendments to the clay showed that CM-carnitine had improved binding to BPA but had a negligible effect for the binding to DBP.[76]

Interactions of Perfluoroalkoxy Alkanes with Parent and Carnitine-Amended Ca^{2+} Montmorillonite Clays in Acidic Conditions

Per- and polyfluoralkyl substances (PFAS) are man-made compounds with widespread use in common consumer and industrial products. While there is limited research on the potential health effects of the bioaccumulation of certain PFAS,[411] current evidence suggests it may cause serious health conditions.[412-414] Among PFAS, PFOA (perfluorooctanoate) and PFOS (perfluorooctanesulfonate) have been widely found in sediment, sludge, municipal wastewater, coastal water, and even tap water.[411] To decrease the use of the highly persistent and cumulative PFOA and PFOS, short-chain PFAS congeners, including GenX (hexafluoropropylene oxide) and PFBS (perfluorobutane sulfonic acid) were introduced as alternatives. However, the hydrophobic chain and hydrophilic functional groups allow PFAS residues to be easily transported in an aquatic environment and adsorbed onto the surfaces of environmental solid matrices.[415] In this study, the binding mechanisms of PFOS, PFOA, PFBS, and GenX to both CM and CM-carnitine in water at pH 2 (simulating stomach conditions) was investigated through MD simulations, structural analysis.[78]

According to both MD simulations and experiments, CM and CM-carnitine were more effective at binding PFOA and PFOS than PFBS and GenX. The higher propensity

of PFOA and PFOS binding to both CM and CM-carnitine than PFBS and GenX was suggested to be due to the longer fluorinated carbon chains of PFOA and PFOS. Both PFOA and PFOS predominantly bound to CM through their hydrophobic tails (Figure V-3A,B). Additionally, in the simulations of CM-carnitine, both PFOA and PFOS predominantly formed hydrogen bonds between their carboxyl group and the carboxyl or hydroxyl group of a bound carnitine with a portion of their fluorinated carbon chains frequently bound directly to the clay surface simultaneously (Figure V-3C,D).

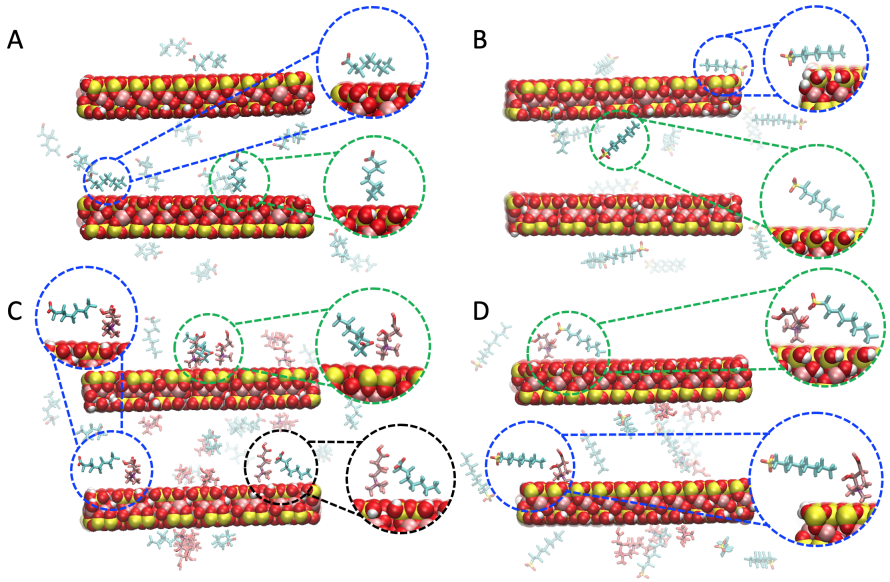

Figure V-3. Molecular graphics of the simulation snapshots containing representative and prominent binding modes for (A) PFOA and (B) PFOS binding to CM, and (C) PFOA and (D) PFOS binding to CM-carnitine. Zoomed in images of the prominent binding modes are encircled by dotted lines and reoriented to facilitate the comparison of different binding modes. This figure is adapted from ref. [78*].

---

* Reprinted with permission from "Enhanced adsorption of per- and polyfluoroalkyl substances (PFAS) by edible, nutrient-amended montmorillonite clays" by Meichen Wang, Asuka A. Orr, Joseph M. Jakubowski, Kelsea E. Bird, Colleen M. Casey, Sara E. Hearon, Phanourios Tamamis, and Timothy D. Phillips, 2021. *Water Research*, 188, 116534, Copyright 2020 by Elsevier.

Combining Experimental Isotherms, Minimalistic Simulations, and a Model to Understand and Predict Chemical Adsorption onto Montmorillonite Clays

The studies of the previous sections provided insight into the molecular mechanisms of specific pesticides, bisphenols, and PFAS binding to parent and nutrient-amended clays with reasonably high qualitative correlations between computations and experiments.[76-78] However, a computationally low-cost, efficient, and accurate method for predicting a compound's absolute free energy of adsorption for the clay and screening which toxic compounds could be mitigated by the clay remained lacking. Additionally, while previously developed data-driven tools aiming to predict adsorption activity of broad-acing materials via regression and dimensionality reduction techniques Qmax and logP values from experiments were able to determine what experimental conditions could maximize adsorption activity,[416] black box methods alone were unable to predict adsorption free energy of compounds to adsorbents due to a limited pool of data.[416]

In collaboration with Dr. Phillips' and Dr. Pistikopoulos' labs, a diverse set of toxic compounds was investigated using minimalistic (one-layered CM) simulations and the hypothesis that their interaction energies with CM, derived from the simulations, as well as intrinsic properties of the compounds and solvents can be used as inputs to a minimalistic model to predict adsorption free energies in agreement with experiments was examined.[79] Simulations of CM in the presence of 35 different toxic compound – solvent systems, for which experimental adsorption free energy values were determined were performed. Of the 35 different toxic compound – solvent systems, experimentally derived adsorption free energies of 24 toxic compound – solvent systems for CM were reported in

previous studies by Dr. Phillip's lab,[76, 77, 319, 320, 323, 324, 417, 418] and experimentally derived adsorption free energies of 11 toxic compound – solvent systems for CM were determined as part of this section by Dr. Phillip's lab.[79] The investigated compounds represent several classes of compounds including bisphenols (BPA, BPS, and BPF), industry solvents (benzene, toluene and phenol), polycyclic aromatic hydrocarbons (naphthalene, benz[e]acephenanthrylene, benzo[a]pyrene, and pyrene), polychlorinated biphenyls (PCB77, PCB126, PCB153, PCB154, PCB155, and PCB157), herbicides (linuron, trifluralin, 2,4-dichlorophenoxyacetic acid, glyphosate, aminomethylphosphonic acid, 2,4,6-trichlorophenol, and paraquat), pesticides (lindane, dieldrin, clofenotane, diazinon, chlorpyrifos, and aldicarb), and mycotoxins (vomitoxin, fumonisin-B1, aflatoxin-B1, and zearalenone).[79]

Overall the simulations were considered minimalistic due to the relatively small size of the simulated system (one-layered CM), the small number of copies of individual toxic compounds (3 copies), and relatively short simulation duration (30 ns) used.[79] Computational efficiency was considered an important factor in this study, as an efficient strategy could enable such approaches to be applied in the future to computationally study, screen, and predict adsorption free energies of multiple individual types of toxic compounds (see minimalistic model below). Upon completion of the minimalistic simulations of CM in the presence of the toxic compounds, the average interaction energy of the toxic compounds, decomposed into van der Waals and electrostatic contributions, were calculated and stored for subsequent use as inputs, in addition to the intrinsic

properties of the compound (charge) and solvent (dielectric constant) to the minimalistic model described in the following passage.[79]

A minimalistic model that can predict adsorption free energies of toxic compounds for CM was developed from several attempts at building a model. The presented minimalistic model is, to our knowledge, the simplest and most accurate model based on the studied set of toxic compounds, based on additional attempts that were performed. In the minimalistic model, the predicted affinity (ΔG^{Pred}) of the compound for CM is estimated from MD-averaged van der Waals (E^{VdW}) interaction energies and the MD-averaged electrostatic (E^{Elec}) interaction energies, the net charge (C) of the compound, and the corresponding dielectric constant (ε) of the solvent.[79] The interaction energy describes the strength of interaction between two molecules (in vacuum) and does not account for polar interaction solvation energy. Thus, it was considered beneficial to divide the electrostatic interaction energy component by the dielectric constant to normalize the term across the different solvents. The division of the electrostatic interaction energy term by the product of the dielectric constant and the net charge also normalized the strength of electrostatic interaction energy for positively charged compounds of different net charge (ranging from 0-2). For the cases in which the net charge of the compound was 0, a "+1" term was also included to avoid division by 0. The model is shown below in Equation V-1:[79]

$$\Delta G^{\text{Pred}} = \alpha\left(E^{\text{VdW}}\right) + \beta\left(\frac{E^{\text{Elec}}}{(\varepsilon)(C+1)}\right) \qquad \text{Equation V-1}$$

In Equation V-1, α and β are tunable parameters. It is worth noting that the model can be considered minimalistic given its simplicity, particularly the fact that only two tunable parameters in combination with four terms (E^{VdW}, E^{elec}, ε, C) are used to capture all needed information; both tunable parameters are unitless as each energy term have units of energy (kcal·mol^{-1}). Other models formulated and tested lacking either the ε, C, or the combination of the two terms performed worse than the presented model in terms of both the R^2 and Root Mean Square Error (RMSE) of the predicted adsorption free energies to the experimental adsorption free energies. Comparison of the predicted and experimentally derived affinities of the toxic compounds to the clay showed that the model provided adequately accurate predicted affinities (Figure V-4). The values of $\alpha = 1.01$ and $\beta = 0.48$ used to calculate the predicted affinities correspond to the average of their values that minimize the model's sum of squared error in each of the 10,000 bootstrap runs.[79]

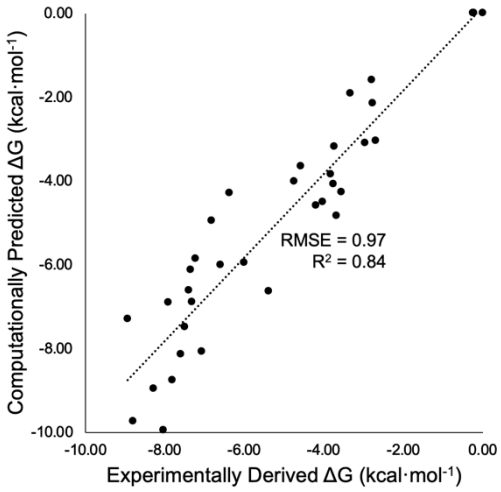

Figure V-4. Correlation of the experimentally observed adsorption free energies (x-axis) and the computationally predicted adsorption free energies (y-axis) of compounds binding to CM (α = 1.01 and β = 0.48).

Additionally, using the minimalistic simulations in conjunction with in-house structural analysis programs,[78] mechanistic insights into the binding of toxic compounds onto CM were derived.[79] According to the simulations, the toxic compounds investigated in this study that bind tightly to CM have one or more of the following characteristics: (1) they contain positively charged groups, (2) they contain a phosphine oxide group, (3) they are halide-rich, (4) they contain hydrogen bond donors/acceptors, and (5) they are large and rigid.[79] Notably, these underlying characteristics associated with CM binding were observed regardless of the use of different solvents in the experiments and in the corresponding simulations.[79] Examples of compounds encompassing these characteristics are shown in Figure V-5.

For the positively charged toxic compounds investigated in this study, based on interaction energy calculations, the positively charged toxic compounds predominantly

bind through electrostatic interactions with the negatively charged interlayer basal surfaces of CM.[79] These strong electrostatic interactions were additionally compounded by strong van der Waals interactions due to the proximity of the toxic compounds to the clay surface.[79]

For toxic compounds containing a phosphine oxide group in this study, both electrostatic and van der Waals interactions contribute to their binding to CM, and all the studied phosphine oxide containing compounds bound to CM through their phosphine oxide groups in their predominant binding modes.[79] This may be due to the fact that phosphorus in phosphine oxide groups is partially positively charged,[195, 419] enhancing the attraction of the compounds to the negatively charged interlayer basal surfaces of CM.[420]

For halocarbon toxic compounds that are halogen-rich, van der Waals interactions predominantly contribute to the binding of halocarbon compounds to CM.[79] Halocarbon toxic compounds with halogen atoms arranged such that multiple halogen atoms can interact with CM simultaneously and maximize molecular contact between the toxic compound and the CM surface generally acquired the highest affinities.[79]

For toxic compounds containing hydrogen bond donor/acceptor groups in this study, both electrostatic and van der Waals interactions contributed to the binding of the toxic compounds to CM. The presence of a hydrogen bond donor allowed these toxic compounds to form hydrogen bonds with the oxygens of the CM interlayer basal surfaces as well as to CM edge surfaces. Additionally, the presence of hydrogen bond donors and acceptors allowed for the formation of solvent-mediated interactions between the

compounds and CM. The presence of hydrogen bond acceptors also allowed toxic compounds to form favorable interactions with Ca^{2+} exchange cations of CM.

Lastly, for large and rigid toxic compounds in this study, van der Waals interactions predominantly contributed to the binding of these inert toxic compounds. These toxic compounds primarily bound to the clay such that their rings are approximately perpendicular to the clay interlayer surface. In their primary mode of binding, the large and rigid toxic compounds appear to be sterically hindered from escaping the interlayer despite their inert chemical properties. For example, benzo[a]pyrene, with a maximum distance across its aromatic rings of approximately 11 Å, occupies nearly the entire interlayer space, which is 15 Å within the simulation due to the periodic boundaries used (Figure V-5, encircled in black). Contrarily, benzene's smaller size appears to allow it to freely flow between clay layers and escape the interlayer.

Overall, the toxic compounds in this study that were chemically inert and relatively small, such as benzene and toluene, did not bind strongly, or minimally bind to CM. Additionally, rigid, but not large, toxic compounds that were sterically hindered from lying coplanar with the interlayer surfaces also minimally bind to CM, such as PCB154. The steric hinderance prevented the formation of dry interfaces between the binding compound and the interlayer surface of CM.

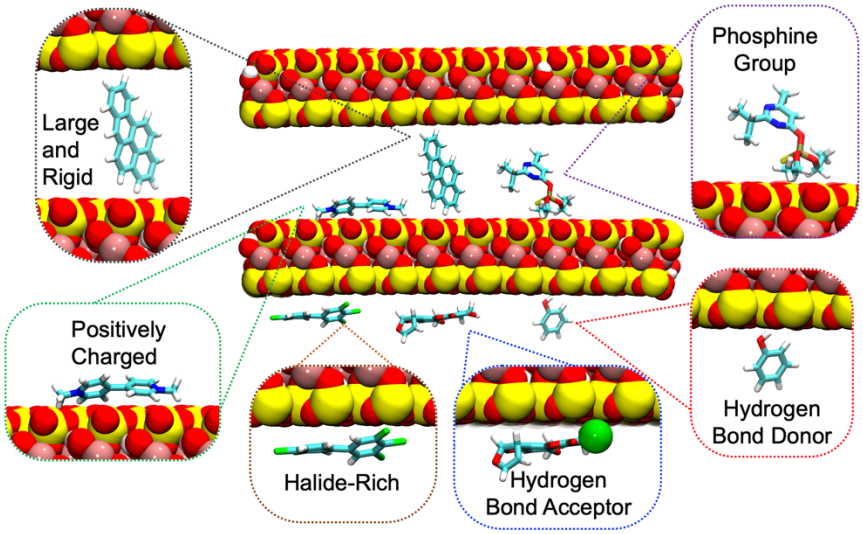

Figure V-5. Molecular graphics image of chemicals encompassing characteristics linked to CM binding.

Given computational resources for relatively short, minimalistic MD simulations and simple energy calculations, the use of systematically executed MD simulations of the one-layer system combined with simple energy calculations and the minimalistic model can save both experimental and computational effort and cost to estimate a compound's affinity for an adsorbent material. Thus, the minimalistic simulations and model can potentially be used to scan hundreds of toxic compounds for their potential to be strongly bound by CM. Notably, the model does not discriminate between toxic and non-toxic compounds; it can determine the binding properties of a compound for CM regardless of its toxicity. Thus, the model could also potentially be used to examine the possibility of other, benign, compounds to be used as amending molecules to produce molecule-amended CM (such as CM-carnitine or CM-choline) with higher affinity for toxic compounds.

Concluding Remarks for Simulating and Analyzing the Binding of Compounds onto Montmorillonite Clays

Montmorillonite clays have been used as therapeutic agents since ancient times.[299-301] Today, they are becoming increasingly attractive candidates for the mitigation of toxic chemicals.[314, 319, 341, 418] Experimental and computational methods have been used to gain valuable insights into the adsorption of compounds onto montmorillonite clays. Experimental adsorption isotherms have provided insights into adsorption pathways and estimates of the thermodynamic properties of various compounds adsorbing onto clay surfaces.[76-79, 320, 323, 324, 417, 421] MD simulations have allowed the dynamic study of chemical compounds binding onto inorganic surfaces, including clays, in atomistic detail and gave insights into the driving forces of adsorption.[37, 77-79, 347, 348, 351, 352, 422-424] In this chapter, the developed clay models and simulation setup have been successfully used in several studies in conjunction with structural and energetic analysis.[76-79] The focus of this chapter was on the simulation setup, which has been generalized such that programs take the selection of adsorbent material, amending compounds (or lack thereof), solvent, toxic compound under investigation, and system size as user input to generate the appropriate simulation files. The systematic generation of simulation systems has allowed the comparison of diverse compounds' propensity to bind to clay[76-78] and ultimately allowed for the development of a minimalistic model that can predict the adsorption free energy of diverse toxic compounds for CM.[79]

CHAPTER VI

PEPTIDE-BASED SELF- AND CO-ASSEMBLY

Elucidating the Pathways of Peptide Self- and Co-Assembly

Introduction

Self- and co-assembly is a process in which individual molecular building blocks spontaneously aggregate form ordered supramolecular structures under mild conditions.[425] In this chapter, self-assembly refers to identical copies of molecular building blocks form supramolecular structures composed purely of the identical building blocks; in co-assembly copies of two or more distinct building blocks form supramolecular structures composed of a mixture of the different building blocks. These ordered structures are generally assembled through noncovalent interactions,[426, 427] and although individual noncovalent interactions are very weak, the combination of several noncovalent interactions together can generate very stable and well-organized structures.[426] Self-assembling proteins and peptides are increasingly gaining interest for engineering functional biomaterials.[428] Their properties can be easily tuned through changes at the sequence level,[22, 23, 429-435] and they can be produced in large quantities through chemical synthesis or recombinant technologies.[436]

Naturally occurring β-sheet forming motifs can be extracted from amyloidogenic proteins or fibrous proteins to yield shorter, self-assembling peptides. For example, short fragments such as KLVFFAE[437] extracted from Aβ (commonly associated with AD) as well as NFGAIL[438] extracted from IAPP (commonly associated with T2D) form ordered,

β-sheet structures. Furthermore, the β-sheet structures formed by these short fragments also possess similar ultrastructural, molecular conformational and cytotoxic properties to the β-sheet structures formed by the full-length, corresponding polypeptides.[437, 438] Short peptides, with sequences extracted from amyloidogenic proteins or fibrous proteins, have been "decorated" with chemically or biologically active groups or motifs to yield functional self-assembled biomaterials. Examples include the A208 peptide containing the IKVAV motif extracted from mouse laminin α1 chain conjugated with RGD tripeptide to promote cell-adhesion on amyloid gel scaffolds,[439] peptides containing the GAITIG motif extracted from adenovirus fiber shaft or GAIIG motif extracted from Aβ with terminal residues mutated to functionalize them for ion/compound binding[22, 23, 430, 440, 441] and cell penetration,[431-433] and peptides containing the VQIVYK motif extracted from tau with mutations introduced for carbon capture.[442, 443]

Interestingly, motifs such as KLVFFAE, NFGAIL, IKVAV, GAITIG, GAIIG, and VQIVYK that confer self-assembly properties contain a high occurrence of aromatic or hydrophobic residues. Because of this, it was hypothesized that hydrophobic interactions, especially interactions between aromatic moieties, may play a central role in amyloid β-sheet formation.[444] While aromaticity is not an essential factor in amyloid self-assembly, for which β-sheet formation is essential, it appeared to accelerate the process for particular peptides.[445] The acceleration of self-assembly was suggested to be due to aromatic side chains having constrained conformations compared to other residues, thereby providing order and directionality as well as minimizing the energy of the amyloid structure.[444]

In addition to aromaticity, hydrophobicity, and polar/electrostatic interactions, ionic interactions can also drive peptide self-assembly. To maximize the formation of as many ionic electrostatic interactions between self-assembled peptides as possible, synthetic ionic-complementary peptides have been rationally designed. Inspired by the Z-DNA binding protein, the first ionic-complementary self-assembling peptide, EAK16 (NH_3^+-AEAEAKAKAEAEAKAK-CO_2^-), was designed[446] with applications in 3D cell culture, tissue engineering, regenerative medicine, and sensory devices.[425, 447, 448] Due to the combination of ionic and hydrogen bond interactions inside the β-sheet structures formed by ionic-complementary peptides, these self-assembled structures are stable under a wide range of temperatures, pH values, and in the presence of denaturing chemicals.[425] Additionally, due to the non-specific nature of hydrophobic interactions, these self-assembled structures may diffuse to minimize equilibrium energy.[449, 450] Among these peptides, the RADA16-I peptide was shown to promote cell growth and tissue regenerations and has been commercialized as PuraMatrix.[449, 451]

As interest in self-assembly systems has increased, so has the quest for the identification of the minimal essential properties needed for self-assembly that could be used for technological applications that rely on mimicking natural self-assembly processes.[445] Very short self-assembling peptides and their mimetics are of special interest as they are easily scalable and can lead to insights into the mechanisms of self-assembly in larger molecules.[452] Following a minimalistic approach to discover the shortest peptide capable of self-assembly, the diphenylalanine (FF) motif, extracted from Aβ, was reported to self-assemble into stiff nanotubes with numerous applications.[453-456] Due to its chemical

simplicity and versatility, FF became a paradigm of study for peptide self-assembly and for designing FF analogs. Materials made from FF with targeted mutations or chemical modifications can be easily fabricated and self-assemble into diverse nanostructures with applications as hydrogels[457-460] and biosensors.[461] Combining one such FF analog, fluorenylmethyloxycaronbyl-diphenylalanine (Fmoc-FF) with other materials either as hybrid hydrogels or through co-assembly resulted in hydrogels with improved properties and functionality.[434, 435, 462, 463] The self- and co-assembly of another FF analog, fluorenylmethyloxycaronbyl-phenylalanine (Fmoc-F) and the functionality of the resulting material can easily be enhanced though incorporating chemical or biological "decorations", such as including halogens on the benzyl side chain.[464-468]

Interestingly, FF was also observed to self-assemble when deposited on a surface by physical vapor deposition, forming arrays of linearly aligned nanotubes called nanoforests.[453] Tuning the deposition time allowed for controllable fabrication of FF nanoforests of a desired thickness, which could be used to produce interfaces with high surface area due to decoration with dense and homogeneous layers of FF nanotubes.[453] It was later suggested that the nanoforests were not entirely composed of FF nanotubes, rather 80-90% of the cylindrical assemblies appeared to be nanorods composed of a modified FF peptide.[469] These modified FF peptides were found to be cyclo-FF peptides (cyclic dipeptide with two phenylalanine side-chains), formed by the dehydration of the linear lyophilized FF peptides during the vapor deposition process.[469] The nanorods formed by cyclo-FF are solid, lacking hydrophilic channels, and are extremely hydrophobic.[469] In line with these results, it was later found that the presence of water

vapor during the deposition process increases the fraction of FF nanotubes to cyclo-FF nanorods in the nanoforest.[470]

Further investigation of self-assembled cyclo-FF structures showed that they possess photoluminescent properties with potential optical applications including light emitting diodes and optical biomarkers.[471, 472] Previous studies have suggested that aromatic residues impart near-ultraviolet (UV) photoluminescence to proteins[473-475] and that cyclo-dipeptides with backbones of 2,5-diketopiperazine configurations, derived from dehydration condensation of linear dipeptides,[476, 477] self-assemble into photoluminescent nanostructures different from their linear counterparts.[478, 479] Additionally, cyclic peptides can be advantageous over linear peptides owing to their conformational stability and decreased conformational entropy in the unfolded state.[480, 481] Thus, a series of aromatic cyclo-dipeptide self-assembles were investigated to evaluate their photoluminescent properties and potential as bio-inspired light-emitting diodes and for *in vivo* bio-imaging.[81, 82, 482]

To enhance the photoluminescence properties of the self-assembled cyclo-dipeptide nanostructures, different peptide sequences, solutions, and ions inclusion were examined. Interestingly, a green fluorescent protein mutant (BFPms1) coordinates with Zn^{2+} via through histidine, resulting in increased fluorescence intensity upon Zn^{2+} binding.[483] According to solid-state NMR spectroscopy and theoretical DFT calculations, histidine coordination with Zn^{2+} is minimal in acidic conditions and its favorability increases as pH levels increase to pH 7.5.[484] Further increasing the pH was suggested to eventually fully deprotonate histidine leading to the formation of a macromolecule

composed of different imidazole rings linked to each other by Zn^{2+}.[484] Inspired by these observations, the self-assembly of cyclo-dihistidine (cyclo-HH) in the presence of Zn^{2+} was investigated. The resulting cyclo-HH : Zn^{2+} assemblies achieved high quantum yields for green fluorescence, exemplifying the potential of such structures to serve as bioinspired, supramolecular alternatives for eco-friendly optoelectronics and bioimaging.[81, 82] The enhanced fluorescence of cyclo-HH : Zn^{2+} gives promise to cyclo-HH as a potential platform for further therapeutic or diagnostic applications. Furthermore, later studies showed that cyclo-HH-ZnI_2 nanowires have hydrolytic activity, with its activity retained after five cycles, making it more robust and durable than other biomolecular artificial hydrolase complexes.[83]

In this chapter, the methodology and applications of MD simulations and a suite of programs performing structural analysis and "thought" free energy calculations developed as part of this doctoral study are described. Fundamental insights into the initial moments of short peptide self- and co-assembly into ordered structures was investigated through the MD simulations combined with the analysis programs. The computational protocol based on MD simulations, structural and energetic analysis for the investigation of the early stages of self- and co-assembly can be considered as generalizable, and thus, may be applied to the investigation of other self- or co-assembly systems involving other molecular building blocks (peptides, molecules, and/or ions) beyond those described in the applications section.

Methodology for Elucidating the Pathways of Short Peptide Self- and Co-Assembly

The methodology for elucidating the pathways of short peptide self- and co-assembly can be outlined as the following stages: (1) all-atom MD simulations are performed to sample the initial stages of peptide self- or co-assembly; (2) clusters are detected based on geometric criteria and ordered or amorphous structures are identified within the MD simulations; (3) structural analysis is performed on the detected clusters to determine their composition and geometric properties; (4) energetic analysis is performed on the detected clusters to compare the thermodynamic favorability of different constructed pathways of self- or co-assembly. The workflow of the presented methodology is presented in Figure VI-1. In this section, the methodology for elucidating the pathways of short peptide self-and co-assembly is described. A detailed description of the methodology is available in refs. [81-84].

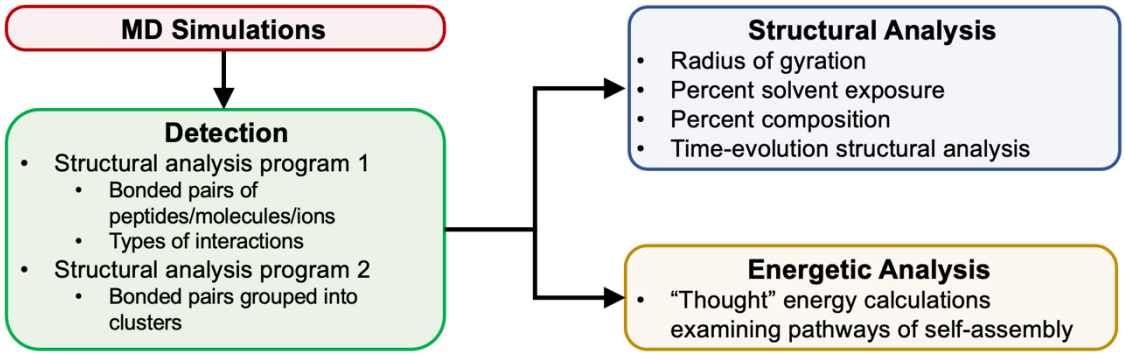

Figure VI-1. Schematic of the overall computational methodology to elucidate the pathways of short peptide self- and co-assembly.

Stage 1: MD simulations of short peptide self-/co-assembly

The self- or co-assembly of short peptides is initially investigated using MD simulations (Figure VI-1, red box). The resulting MD simulation snapshots will subsequently be used as input for structural analysis and energy calculations in the later stages. The initial structures for the MD simulations are built by placing copies of the short peptide under investigation and any co-assembling molecule or ion, if present in the self-assembly system of interest, inside a grid. The conformations and orientation of each copy of the short peptide and co-assembling molecule within the grid are random and generated from short 1 ns simulations of a single molecule at infinite dilution using the GBSW implicit solvent model.[133] In the grid, the short peptides and molecules or ions under investigation are placed such that they are equally spaced such that the initial distance between each neighboring peptide, molecule, and/or ion is within the cutoff of non-bonded interactions. The initial distance is set aiming to facilitate the formation of an initial aggregate such that each peptide, molecule, and/or ion can initially "interact" with a neighboring each peptide, molecule, and/or ion within the simulations. The number of copies of peptides, molecules, and/or ions within the simulation system should be sufficiently large to enhance statistical analysis of the interactions formed within the simulations.[81-84]

The initial configuration of the grid of peptides, molecules, and/or ions under investigation are subsequently solvated in solvent boxes. The composition of the solvent boxes is selected to be in line with experimental conditions. For example, in line with experimental conditions, methanol was used for ref. [81], isopropanol was used for ref. [82, 83],

and water was used for ref. [84]. The size of the solvent box may be constructed such that the concentration of the peptides, molecules, and/or ions is higher than used within experiments to facilitate the interaction of the peptides, molecules, and/or ions as well as to enhance the sampling and formation of clusters within the simulations.[485] In refs. [81-84], the size of the solvent boxes were built to increase the simulated concentration, compared to experiments, of the self- and co-assembling peptides, molecules, and/or ions to facilitate self- and co-assembly.

The solvated simulation systems are simulated with multi-ns MD simulations. Prior to the execution of the production simulation runs, the simulation systems are first subjected to constrained energetic minimizations and short equilibration. The energy minimization and short equilibration aims to allow the solvent molecules to relax around the assembling peptides, molecules, and/or ions. Subsequently, all constraints imposed on the simulation systems are released for the production simulation run. During the production run, simulation snapshots are saved throughout the duration of the simulations for subsequent structural and energetic analysis. The duration of the MD simulations may be tailored to the system under investigation.[81-84]

Stage 2: Detection of Clusters Based on Geometric Criteria

Within the MD simulations the investigated molecular building blocks (peptides, molecules, and/or ions) may self- or co-assemble into clusters. These clusters can be structurally characterized by post-processing the simulation trajectories though in-house structural analysis programs developed as part of this doctoral study. Two structural

analysis programs are used to detect clusters of molecular building blocks, the first (structural analysis program 1) identifies bonded pairs of molecular building blocks, and the second (structural analysis program 2) identifies clusters of molecular building blocks based on the identified bonded pairs.

Structural analysis program 1 detects interacting or bonded pairs of molecular building blocks (peptides, molecules, and/or ions) within each simulation snapshot based on atom-to-atom distances. If the distance between two atoms belonging to different individual molecular building blocks is less than a user defined distance cutoff, then the two atoms are considered to be bonded. If a pair of atoms are considered bonded, then the program compares the interacting atoms to a list of interaction type definitions, and information on how the atoms are bonded is recorded. After structural analysis program 1 is executed, a list of bonded pairs of molecular building blocks (pairs of peptides, molecules, and/or ions), the simulation snapshot in which they are interacting, and the type of interaction through which the two molecular building blocks are bonded is produced.

Subsequently, structural analysis program 2 reads the data in the list of bonded pairs of molecular building blocks (peptide, molecule, or ion) to group the molecular building blocks into clusters such that a number of s molecular building blocks are defined to form a cluster when each molecular building block is bonded to at least one other molecular building block in the cluster. The clustering process operates in two steps: In the first step, temporary clusters of increasing size are detected, and in the second step, a final list of clusters is output with redundancies or the presence of smaller clusters within larger clusters are removed.

For the first step, clusters of increasing size are identified with individual molecular building blocks added one at a time for each simulation snapshot. For each snapshot, temporary clusters of two molecular building blocks are compared to a list of bonded pairs of molecular building blocks. If any one of the molecular building blocks in the list of bonded pairs is present in the temporary cluster of two, then the molecular building block is added to the cluster and the cluster size is expanded to a temporary cluster of three. Next, the temporary clusters of three molecular building blocks are compared to the same list of bonded pairs of molecular building blocks. If any one of the molecular building blocks in the list of bonded pairs is present in the temporary cluster of three, then the molecular building block is added to the cluster and the cluster size is expanded to a temporary cluster of four. This process is repeated until no larger clusters are detected. In this step, temporary clusters of smaller sizes can also be detected within larger clusters. For example, if a given molecular building block belongs to a cluster of size 5, then the molecular building block will also be present in smaller temporary clusters of sizes 4, 3, and 2. These redundancies will be removed in the next step.

In the second step, the presence of smaller clusters within larger clusters are removed. For example, if a given molecular building block is present in a cluster of size 5, then it is no longer present in a cluster of size 4, 3, or 2, and the smaller temporary clusters are removed. In this way, redundant smaller clusters within larger clusters are removed and an individual molecular building block cannot belong to two separate clusters at once. After structural analysis program 2 is executed, a list of clusters, the size of each cluster, the residue identification number (unique to each individual molecular

building block) of the molecular building blocks comprising each cluster, and the simulation snapshot in which the cluster is detected is produced. The clusters of molecular building blocks detected by the in-house structural analysis programs are extracted from the simulation trajectories and further analyzed in the following stages.

Stage 3: Structural Analysis of the Detected Clusters to Determine Composition and Geometric Properties

The geometric properties of the detected clusters extracted from Stage 2 can subsequently be analyzed. Time-evolution structural analysis can be performed to track the formation of clusters within the MD simulations. Additionally, the geometric properties of the formed clusters analyzed in Stage 3 include compactness, and, for co-assembly simulation systems, the cluster composition and the location of peptides, molecules, or ions within the clusters. In the following analysis, the detected clusters in the simulations are isolated from the simulation systems prior to the structural analysis.

Time evolution structural analysis tracking geometric properties with respect to simulation time can provide insights into the process by which the molecular building blocks within the detected clusters self- or co-assemble. The time evolution structural analysis program uses the list of bonded pairs of molecular building blocks and their interaction types generated from the structural analysis program 1 of Stage 2, to produce the total number of instances for each interaction type detected within user-defined discrete time blocks of the simulation. This data is then plotted with respect to simulation time to observe the order by which the interactions are formed. For co-assembly systems,

the time evolution structural analysis program additionally uses the list of clusters generated from structural analysis program 2 to determine the composition of the clusters (percent of each type of molecular building block within the cluster) formed within the simulations. Thus, for each snapshot, the time evolution structural analysis program outputs the average percent composition of each type of molecular building block within clusters that are composed of molecular building blocks that eventually form larger clusters (the cluster size(s) qualifying as large is defined by the user). For example, in refs. [81-83], specific interactions between cyclo-HH in different environments were tracked to uncover the order in which interactions were formed between the pairs of cyclo-HH to ultimately form ordered elementary structures of β-bridge bonded cyclo-HH pairs. In ref. [82], the time evolution composition of clusters composed of individual peptides, molecules, and ions that eventually form large clusters (at least 10 cyclo-HH or epirubicin) was tracked to uncover the process by which the individual peptides, molecules, and ions co-assemble.

The radius of gyration of each detected cluster can provide a metric of the compactness of the clusters. Additionally, the radius of gyration of specific peptides, molecules, or ions within the detected clusters can provide insights into their compactness within the clusters. The radius of gyration structural analysis program reads the list of clusters produced by structural analysis program 2 of Stage 2. For each detected cluster, the radius of gyration structural analysis program uses Wordom[486, 487] to calculate the radius of gyration of the collection of molecular building blocks that make up the cluster. From the radius of gyration structural analysis program, a list of the average radius of

gyration for each cluster size is produced. When comparing the radius of gyration of clusters formed across different simulation systems, as in ref. [84], or the radius of gyration of specific type of molecular building blocks across different simulation systems, as in ref. [82], it is important to ensure a "fair" comparison. For example, in ref. [84], in the calculations comparing the compactness of peptides in clusters formed by Fmoc-3,5F-Phe peptides versus Fmoc-3,4F-Phe peptides, the comparison was only performed between clusters containing the same number of peptides. In ref. [82], in the calculations comparing the compactness of Zn^{2+} ions in clusters across different simulation systems, the comparison was performed between clusters containing the same number of cyclo-HH and the number of Zn^{2+} ions within each cluster of the same size were similar across the two systems. In general, larger relative radius of gyration values indicate lower compactness, or lower packing density of the peptides, molecules, or ions in the clusters, while lower relative radius of gyration values indicates higher compactness, or higher density of peptides, molecules, or ions in the clusters.[488]

For co-assembled clusters containing different types of molecular building blocks the radius of gyration can also indicate the location of the specific types of molecular building blocks in a cluster with respect to other types of molecular building blocks of the same cluster. The radius of gyration of one type of molecular building block within a cluster can be compared to the radius of gyration of another within the same cluster to indicated if one type of molecular building block is encapsulated by the other. In the case of co-assembly, the radius of gyration structural analysis program reads the list of clusters produced by structural analysis program 2 of Stage 2. For each detected cluster and each

type of molecular building block of interest, the radius of gyration structural analysis program uses Wordom to calculate the radius of gyration for molecular building blocks specified by the program.[486, 487] For example, if two molecular building blocks are of interest, then the radius of gyration for each molecular building block per cluster is calculated for the portion of the cluster comprising only one type of molecular building block, with all other molecular building blocks omitted, as well as the remaining type of molecular building blocks that make up the cluster, with the former type of molecular building blocks omitted. From the radius of gyration structural analysis program, a list of the average radius of gyration for each molecular building block of interest per cluster size is produced. In this case, a "fair" comparison is also important when comparing the radius of gyration of specific peptides, molecules, or ions within different clusters of the same simulation system. As in ref. [82], isolating the analysis to clusters of a given percent composition of each type of peptide, molecule, or ion and calculating the difference in radius of gyration between the entities of interest per cluster (rather than comparing the average radius of gyration for each type of peptide, molecule, or ion across all clusters) can enable a "fair comparison". For example, in ref. [82], the calculations were performed for clusters containing at least 10 cyclo-HH or epirubicin, with a composition ranging from 30% cyclo-HH and 70% epirubicin to 70% cyclo-HH and 30% epirubicin. This percent composition criterion was introduced to ensure that each cluster had a sufficient number of cyclo-HH and epirubicin. [82]

The percent solvent exposure of a molecular building block within a cluster can provide insights into the geometric properties of the cluster and the location of each

molecular building block within the cluster. The percent solvent exposure of a molecular building block can be measured by the solvent accessible surface area (SASA) of the molecular building block divided by the total molecular surface area (TSA) of the same molecular building block. The larger the percent solvent exposure of a molecular building block, the more exposed it is to the solvent and the more likely it is to be at the surface of the cluster; the smaller the percent solvent exposure of a molecular building block, the more "buried" it is and the more likely it is to be encapsulated in the interior of the cluster. The solvent exposure structural analysis program reads the list of clusters produced by structural analysis program 2 of Stage 2. For each detected cluster, the SASA and TSA of each molecular building block comprising the cluster is calculated using Wordom[486, 487] and the percent solvent exposure of each molecular building block is calculated. From the solvent exposure structural analysis program, the average percent solvent exposure of each type molecular building block is printed. The use of percent solvent exposure calculations in characterizing the geometric properties of cyclo-HH co-assembled with epirubicin, NO_3^-, and Zn^{2+} is described in ref. [82].

Stage 4: Energetic Analysis of the Detected Clusters to Compare Thermodynamic Favorability of Different Self-Assembly Pathways

The detected clusters extracted from Stage 2 can also subsequently be analyzed through association free energy calculations. Association free energy calculations can provide valuable insights into the mechanism and driving forces leading to the co-assembly and stabilization of clusters formed by the molecular building blocks under

investigation. The energy calculations can also serve to complement the structural analysis and be a measure to ensure that the conclusions from both structural and energetic analyses correlate and are consistent. In the energy calculations, the detected clusters in the simulations are isolated from the simulation systems and undergo a series of "thought" energy calculations. In the series of "thought" energy calculations, the MM–GBSA[49] approximation is used to provide a relatively fast and effective means to evaluate the association free energy of the clusters, and the isolated cluster is subjected to different conditions to examine different potential pathways of co-assembly.

In the in-house energy calculation program developed as part of this doctoral study, the program isolates each cluster detected by structural analysis program 2 of Stage 2 and executes a CHARMM[197] script for each "thought" energy calculation, extracts energetic data from the CHARMM[197] output, and calculates the association free energies normalized by the size of the cluster. Thus, from the energy calculation program, the average energy per cluster size for each "thought" energy calculation, which corresponds to a constructed thermodynamic self-assembly pathway, is produced. The thermodynamic pathway with the lowest average free energy is considered the most energetically favorable self-assembly pathway. Possible "thought" energy calculations that may be performed are detailed in refs. [81, 82]. Insights gained from the different energy calculations can lead to the formulation of additional "thought" energy calculations exploring intermediate states that lead to the final formation of the cluster as in ref. [82]. In the following, example "thought" energy calculations and what they represent are presented:

1. To represent the free energy for isolated individual molecular building blocks, which are part of a given cluster, to spontaneously self-assemble into the cluster, the MM–GBSA association free energy is calculated through Equation VI-1.

$$\Delta G(s)=E_{\text{cluster}}-\sum_{i=1}^{s} E_i \qquad \text{Equation VI-1}$$

The energy of the cluster, E_{cluster}, includes the intra- and intermolecular energies of the constituent molecular building blocks, with all other molecular building blocks deleted. The energies of the isolated, individual molecular building blocks of the cluster, E_i, is calculated by assuming that each molecular building blocks, i, has the same conformation as it does within the cluster, but isolated and fully immersed in solution, with all other molecular building blocks within the cluster deleted.

2. To represent the free energy for isolated individual molecular building blocks, which are part of a cluster, to aggregate onto a preformed portion of the same final cluster, the MM-GBSA association free energy is calculated through Equation VI-2.

$$\Delta G(s)=E_{\text{cluster}}-E_{\text{preformed}}-\sum_{i=1}^{s} E_i \qquad \text{Equation VI-2}$$

The energy of the cluster, E_{cluster}, includes the intra- and intermolecular energies of the constituent molecular building blocks, with all other molecular building blocks deleted. The energy of the cluster, $E_{\text{preformed}}$, includes the intra- and intermolecular energies of the molecular building blocks constituting the preformed portion of the

cluster, with all other molecular building blocks deleted. The energies of the isolated, individual molecular building blocks of the cluster, E_i, are calculated by assuming that each molecular building blocks, i, has the same conformation as it does within the cluster, but isolated and fully immersed in solution, with all other molecular building blocks within the cluster deleted.

3. To represent the free energy for two preformed portions of the cluster to aggregate with each other, the MM-GBSA association free energy is calculated through Equation VI-3.

$$\Delta G(s)=E_{\text{cluster}}-E_{\text{preformed 1}}-E_{\text{preformed 2}} \qquad \text{Equation VI-3}$$

The energy of the cluster, E_{cluster}, includes the intra- and intermolecular energies of the constituent molecular building blocks, with all other molecular building blocks deleted. The energy of the cluster, $E_{\text{preformed 1}}$, includes the intra- and intermolecular energies of the molecular building blocks constituting the first preformed portion of the cluster, with all other molecular building blocks deleted. The energy of the cluster, $E_{\text{preformed 2}}$, includes the intra- and intermolecular energies of the molecular building blocks constituting the second preformed portion of the cluster, with all other molecular building blocks deleted.

In the example energy calculations, the cluster, preformed portions of the cluster, and individual molecular building blocks of the cluster are all assumed to be fully immersed in pure solvent through the deletion of molecular building blocks not involved in the energy calculation. Additional energy calculations can also be performed with the

energy calculation program to examine hypothetical (i.e., thought) pathways in which the interior of the cluster, exterior of the cluster, and/or individual molecular building blocks of the cluster are in a peptide-like environment (given that the clusters are formed with peptides). In such thought calculations, the nonpolar component of the association free energies is calculated as described in Equation VI-1, Equation VI-2, and Equation VI-3. The polar component of the association free energies, however, are calculated by preserving the charge of the molecular building blocks involved in the energy calculation and setting the charge of all molecular building blocks within the cluster to zero. For example, to examine the contribution of molecular building blocks co-assembling in a peptide-like environment, the association free energy would be calculated through Equation VI-2, except the polar component of E_i for a given molecular building block would be calculated by setting the charge of all other molecular building blocks within the cluster to zero.[81, 82] In this way, the calculation represents the energy in a peptide-like dielectric environment, rather than a pure solvent dielectric environment.[81, 82]

Within these calculations, the GBSW implicit solvent model[133] is used to account for the solvent. In the implicit solvent model, the dielectric constant can be tuned in accordance with the solvent used in the simulations and experiments.[81, 82] Additionally, as the inclusion of non-polar solvation in these calculation is important and may cause inaccuracies in the calculated energies if not accounted for carefully, the nonpolar solvation effects may be omitted by the user[81, 82] if comprehensive studies of the appropriate surface tension coefficient for the solvent under investigation have not been performed with the particular implicit solvent in conjunction with the GB model used (e.g.,

GBSW[133]). Nevertheless, supporting calculations may be performed using the default surface tension coefficient value determined for water corresponding to the GBSW implicit solvent model,[133] 0.03 kcal·mol^{-1}, to ensure the overall trends remain the same.

Applications of Elucidating the Pathways of Peptide Self- and Co-Assembly

The methodology to elucidate the pathways of peptide self-assembly presented in the previous section has been used in several studies investigating the self- and co-assembly of cyclo-HH in different environments.[81-83] The methodology was also used to investigate the self-assembly of Fmoc-F derivatives.[84] All experiments for the studies investigating cyclo-HH self-assembly were all performed by members of Dr. Gazit's lab at Tel Aviv University, Tel Aviv, Israel and collaborators of Dr. Gazit's lab. Experiments for the investigation of Fmoc-F derivative self-assembly were all performed by members of Dr. Adler-Abramovich's lab at Tel Aviv University, Tel Aviv, Israel. Contributors to these works are listed in the authors list of refs. [81-84]. Additional details of these study are provided in refs. [81-84].

Elucidating Cyclo-HH Self- and Co-Assembly: Minimalistic Bio-assemblies with Enhanced Fluorescence, Drug Encapsulation, and Catalytic Properties

Bioinspired assembling materials have demonstrated promising potential to serve as the foundation for next-generation photoelectronics.[489-492] Inspired by the molecular structure of a green fluorescent protein mutant, BFPms1,[483] the effect of the inclusion of different ions was explored to manipulate the assembly of cyclo-HH into a peptide material with the high-fluorescence efficiency in collaboration with Dr. Gazit's lab. To study the effect of ions (no ions, $ZnCl_2$, ZnI_2, $Zn(NO_3)_2$) on the self-assembly of cyclo-HH, MD simulations and free energy calculations in conjunction with structural analysis programs were performed.[81-83]

According to MD simulations investigating the initial nucleation stage of cyclo-HH self-assembly in the presence of Zn^{2+} and Cl^-, pairs of cyclo-HH tended to form anti-parallel β-bridge conformations (Figure VI-2A).[81] In contrast, in the absence of Zn^{2+}, pairs of cyclo-HH tended to form parallel β-bridge conformations (Figure VI-2B).[81] Additional MD simulations and analysis also showed that the self-assembly of cyclo-HH in the presence of Zn^{2+} and I^-, instead of Cl^-, also resulted in pairs of cyclo-HH predominantly forming anti-parallel β-bridge conformations.[83] These findings are consistent with X-ray crystallography structures (Figure VI-2A,B) and suggested that Zn^{2+} affects the association of cyclo-HH and the antiparallel β-bridge conformation was stabilized by Zn^{2+} through "locking" of the pair of cyclo-HH.[81, 83] These β-bridge dimeric conformations are the basic units and act as the fundamental building blocks to self-assemble into larger structures, thus implying the structural basis underlying the photoactive properties.[81, 83] According to time-evolution structural analysis and hypothetical or "thought" free energy calculations exploring different self-assembly pathways, the doping of cyclo-HH by Zn^{2+} is initially driven by individual cyclo-HH attracting and pulling of the metal ions from the solvent into the peptide environment, rather than by direct interaction in the solvent, thus forming an "environment-switching" doping mechanism (Figure VI-2C).[81, 83]

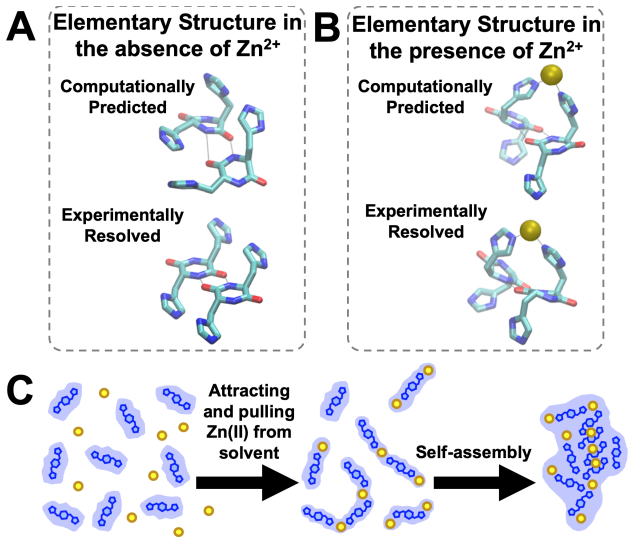

Figure VI-2. Overview of the formation of clusters by cyclo-HH in the absence and presence of Zn^{2+}. (A) Comparison of most prominent elementary structure of cyclo-HH in the absence of Zn^{2+} observed in simulations with experimental crystallography. (B) Comparison of most prominent elementary structure of cyclo-HH in the presence of Zn^{2+} observed in simulations with experimental crystallography. (C) Schematic of the proposed "environment switching" mechanism for cyclo-HH : Zn^{2+} assembly formation.

According to MD simulations investigating the initial nucleation stage of cyclo-HH self-assembly in the presence of $Zn(NO_3)_2$ compared to $ZnCl_2$ in isopropanol, Zn^{2+} ions were more densely packed and concentrated in the clusters formed in the presence of NO_3^- for containing the same number of cyclo-HH.[82] Combined experimental and computational analysis suggested that cyclo-HH : Zn^{2+} is encapsulated into cyclo-HH : NO_3^- assemblies and that immobilization of the peptide oligomers by a high-stiffness scaffold would limit energy dissipation during thermal relaxation pathways for better quantum yield and fluorescence intensity.[82]

Motivated by the enhanced optical properties of the cyclo-HH : $Zn(NO_3)_2$, the capability of the material to serve as an emissive nanocarriers for cancer drug delivery was

investigated. MD simulations of cyclo-HH in the presence of $Zn(NO_3)_2$ and Epirubicin, a cancer drug, showed that Epirubicin and Zn^{2+} ions were primarily located in the interior nucleus whereas cyclo-HH and NO_3^- were primarily at the exterior surface, thereby depicting the "self-encapsulation" properties of the system (Figure VI-3B,D). According to structural and energetic analysis, the cyclo-HH : $Zn(NO_3)_2$: Epirubicin clusters were formed with the Zn^{2+} ions and Epirubicin molecules first aggregating to form the inner nucleus of the clusters followed by the assembly of individual pieces of cyclo-HH and NO_3^- exteriorly wrapping around the preformed interior (Figure VI-3B,C). Further energetic analysis suggested that Zn^{2+} ions and initially formed Epirubicin clusters are in a peptide-like environment prior to the outer cyclo-HH and NO_3^- wrapping around the epirubicin cluster (Figure VI-3A). Experiments confirmed that the "self-encapsulation" properties of cyclo-HH:$Zn(NO_3)_2$ can be used to enhance the delivery of Epirubicin to HeLa cells and can be used as an advanced nanocarrier with integrated *in situ* monitoring.[82]

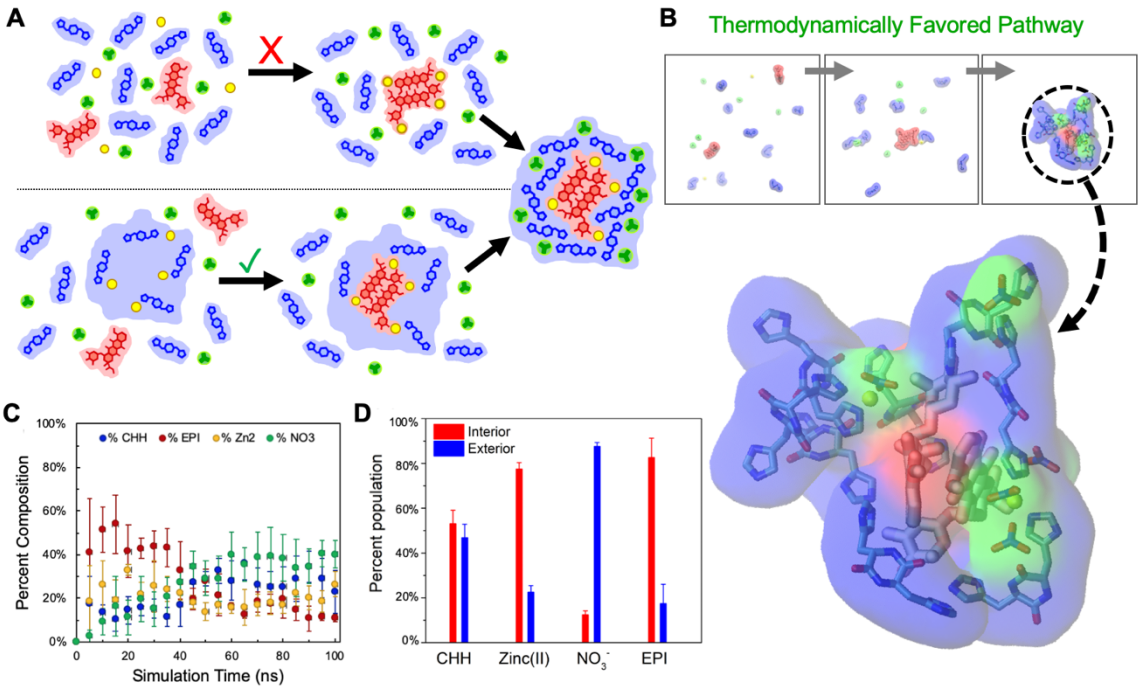

Figure VI-3. Overview of the formation of clusters by cyclo-HH in the presence of $Zn(NO_3)_2$ and Epirubicin. (A) Schematic of potential pathways for the "self-encapsulation" of Epirubicin by cyclo-HH: $Zn(NO_3)_2$. The free energy to associate the interior cluster (Epirubicin and Zn^{2+}) from the individual entities is unfavorable (indicated by the red 'X'). However, the free energy to associate the interior cluster (Epirubicin and Zn^{2+}) is favorable if the Zn^{2+} ions and Epirubicin are in a peptide-like environment prior to association (indicated by the green check mark). (B) Epirubicin encased by cyclo-HH:$Zn(NO_3)_2$ observed in MD simulations. (C) The time evolution percent composition was calculated for clusters composed of the entities (cyclo-HH, Zn^{2+}, NO_3^- or Epirubicin) that eventually form clusters containing at least 10 molecules with a composition ranging from 53% cyclo-HH and 83% Epirubicin to 47% cyclo-HH and 17% Epirubicin. (D) Percent population of cyclo-HH, Epirubicin, Zn^{2+} and NO_3^- within the interior and exterior of the clusters containing at least 10 molecules with a composition ranging from 53% cyclo-HH and 83% Epirubicin to 47% cyclo-HH and 17% Epirubicin. This figure is an adaptation from ref. [82].

Differentiating the Self-Assembly of Fmoc-Phe Derivatives

Fmoc-phenylalanine (Fmoc-Phe) based materials are also attractive due to the wide variety of unique properties that can be obtained through chemical "decorations".[464-466, 493-496] In the present study the methodology used to elucidate the pathways of cyclo-HH self-assembly in the previous section was used to investigate the self-assembly of two Fmoc-Phe derivatives, Fmoc-3,4-difluoro-phenylalanine (Fmoc-3,4F-Phe) and Fmoc-3,5-difluoro-phenylalanine (Fmoc-3,5F-Phe), in collaboration with Dr. Adler-Abramovich's lab.[84] To study and compare the properties of Fmoc-3,4F-Phe and Fmoc-3,5F-Phe within their first moments of self-assembly, MD simulations and structural analysis for each Fmoc-Phe derivative were performed.[84]

Within MD simulations, Fmoc-3,4F-Phe and Fmoc-3,5F-Phe peptides both self-assembled into aggregates significantly stabilized by π-π interactions between aromatic groups as well as F-F/F-Phe contacts and hydrogen bonds between terminal carboxyl groups or backbone amide and carbonyl groups of peptides. While both Fmoc-3,4F-Phe and Fmoc-3,5F-Phe were observed to form structures reminiscent of antiparallel β-sheet structures at a similar rate, Fmoc-3,4F-Phe more frequently formed structures reminiscent of parallel rather than antiparallel β-sheet structures (Figure VI-4A). Importantly, this ordered structure was reminiscent of the crystal structure of self-assembled Fmoc-3,4F-Phe. Additionally, Fmoc-3,4F-Phe more frequently formed face to face π-π interactions between Fmoc-Phe and Phe-Phe groups (Figure VI-4A). These interactions could potentially enable the formation of the parallel β-sheet-like structures observed in the MD simulations and crystal structure (Figure VI-4A,B boxed in red dotted lines). Contrarily,

Fmoc-3,5F-Phe aggregates were more frequently stabilized by Fmoc-Fmoc π stacking interactions, contacts between F of the 3,5F-Phe group and Fmoc, in the absence of Fmoc-Phe π stacking (Figure VI-4B boxed in red, black, and blue dotted lines respectively). In addition to the differences in the frequency of interactions formed within their aggregates, the calculated radius of gyration of the derivatives within the clusters formed by Fmoc-3,5F-Phe was also consistently lower compared to those formed by Fmoc-3,4F-Phe across aggregates of different sizes (Figure VI-4C). Thus, for clusters containing the same number of peptides, clusters formed by Fmoc-3,5F-Phe were more densely packed, in line with experimental density measurements.[84] This study is currently under review.[84]

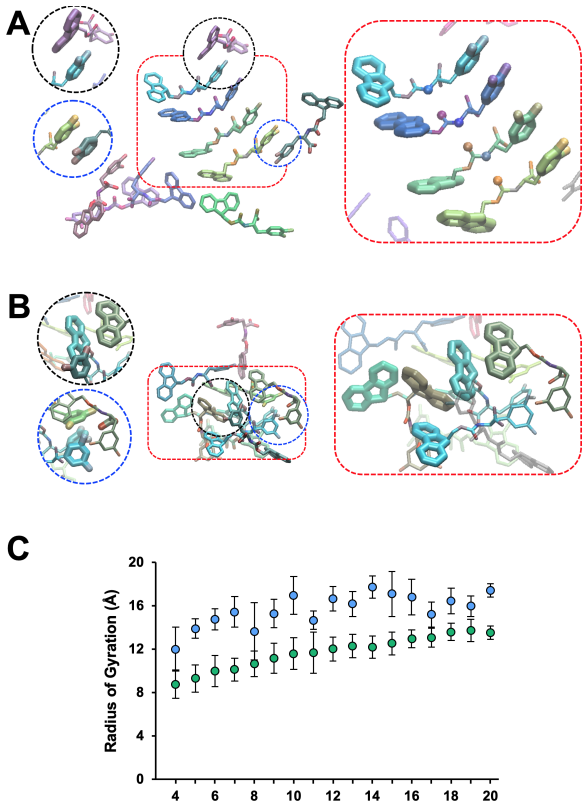

Figure VI-4. Structural properties of the clusters formed by Fmoc-3,4F-Phe and Fmoc-3,5F-Phe peptides within MD simulations. Representative aggregate of 10 (A) Fmoc-3,4F-Phe and (B) Fmoc-3,5F-Phe peptides with key differences in interactions encircled in dotted lines and zoomed in. (C) Radius of gyration (Å) of peptides within the clusters observed in MD simulations of Fmoc-3,4F-Phe (blue) and Fmoc-3,5F-Phe (green) peptides.

Concluding Remarks for Elucidating the Pathways of Short Peptide Self- and Co-Assembly

Several minimalistic peptide self- and co-assembly have been precisely designed to encompass desired functionalities and properties (reviewed in ref. [497]). The MD simulation setup as well as the structural analysis and free energy calculation programs that were developed for this chapter allow for the detailed computational investigation of the initial stages of short peptide self- and co-assembly. The simulations and analysis programs have been successfully used to study the initial stages of self- and co-assembly of cyclo-HH.[81-83] The MD simulation setup has also been used for the study of other peptide-based self-assembly beyond cyclo-HH.[84] The generalizability of the developed simulation setup, structural analysis programs, and energy calculation programs may be applied in the investigation of self- and co-assembly involving other short peptides.

CHAPTER VII

SUMMARY

Current and Future Perspectives

The overarching theme of this doctoral study is the development of different computational tools that can serve as bridges to solve challenging problems that are otherwise difficult to solve using existing experimental or computational approaches. The computational tools developed in this doctoral study aided in the solution of challenging problems and have laid the foundation for future studies involving particular problems within the areas of 1) small-molecule ligands binding to proteins, 2) RNAs binding to proteins, 3) inhibition of amyloid self-assembly, 4) ligands binding to adsorbents, and 5) peptide-based self-assembly and co-assembly. In what follows, the current and future perspectives of each computational tool is described.

Docking-Refinement Protocol for Small-Molecule Ligand : Protein Complexes

The determination of the orientation and conformation of structurally and physicochemically similar small-molecule ligands in ligand : protein complexes experimentally can be highly challenging.[51, 52] Additionally, conventional molecular docking methods face challenges in accounting for the conformational flexibility in the protein receptor[498] as well as inaccuracies in the calculations of binding energies.[53, 499] This limits their ability to accurately predict and differentiate the binding of physicochemically similar small-molecule ligands.[53]

To overcome these challenges and understand how subtle structural and physicochemical differences in ligands can lead to functional differences upon binding to a protein receptor, a novel docking-refinement protocol was developed. It was applied delineate the binding of similar ligands. The docking-refinement protocol has been applied to elucidate the binding of small-molecule ligands binding to human and mouse AhR,[56-58] human COUP-TF1 and COUP-TF2,[59] as well as *E. coli* Tsr.[54, 55] The resulting binding modes, MD simulations, and free energy calculations have suggested key interactions for the ligands' binding.[54-59] Furthermore, comparisons between the key interactions of structurally and physicochemically similar ligands with different functional properties have also indicated potential interaction switches for their biological activity;[54-59] the use of binding modes derived from the docking-refinement protocol combined with multi-ns MD simulations predicted the stronger chemoattractant strength of one enantiomer over another, with experiments validating the computational prediction.[55]

The use of the docking-refinement protocol may not necessarily be limited to the comparison between similar ligands. For example, the docking-refinement protocol was applied to refine the binding mode of a virtually screened compound with demonstrated binding to complement C3c.[500] Uncertainties in the binding mode of small-molecule ligands has major implications for drug discovery. Although specific drug functional groups may be identified, the precise binding mode of each functional group may be uncertain, impacting the subsequent pharmacophore models used to identify additional drug candidates.[51] Nevertheless, for its application in drug discovery and virtual screening, the efficiency of the protocol must be improved without impairing its accuracy.

The investment in added computational cost to implement the docking-refinement protocol has proven useful in differentiating the binding modes of structurally and physicochemically similar ligands with different biological activities.[54-59] Future improvements to the docking-refinement protocol by members of the Tamamis Lab include enhancing its efficiency without detriment to its accuracy, which is among the goals of the Tamamis Lab.

Protocol for the Characterization of Modified RNA : Protein Interactions

The coordination of RNAs with RNA binding proteins is crucial for many vital biological processes.[2-6, 8, 146] The presence of chemical modifications to canonical RNAs can influence the functions of RNA : protein interactions.[190] Limitations in current methods have limited progress in mapping the epitranscriptome.[62-64] Thus, despite the fact that over 150 chemically modified RNAs have been identified,[60] their effect on the function of RNA : protein complexes has been largely unexplored.[61]

To overcome the lack of methods to uncover the repertoire of modified RNAs that can be recognized by a target protein, a protocol for the characterization of modified RNA : protein interactions was developed in which a suite of programs screen and detect modified RNAs prone to interact favorably with a target protein. The protocol for the characterization of modified RNA : protein interactions has been applied to map the modified RNAs recognized by RNA binding proteins *E. coli* PNPase,[65, 67] human YTHDF1,[67] human NOVA1,[67] and human TDP-43.[67] An analogous version of the

protocol has also been used in an "inverse" fashion to design PNPase mutants with enhanced affinity for an oxidized modified RNA (8-oxoG).[66]

The protocol for the characterization of modified RNA : protein interactions has also been recently used as a steppingstone and basis for a preliminary protocol for the characterization of modified DNA : protein interactions. As in the protocol developed for modified RNAs, the preliminary protocol for the characterization of modified DNA : protein interactions uses an initial structure of the DNA : protein complex structure and a library of DNA modifications under investigation to identify DNA modifications prone to interact favorably with the target protein. The library of modifications for the protocol for the characterization of modified DNA : protein interactions corresponds to all 42 experimentally verified DNA modifications,[501] are preliminarily organized into trees based on their structural and physicochemical properties, starting from the least complex chemical modifications at the base, and ending at the most complex modifications. Of the 42 DNA modifications, 13 have been parametrized by MacKerell and Nilsson groups,[177] and 29 were parametrized through CGenFF.[193-196] To account for base-pairings in the preliminary version of the protocol, the complementary base pair to the DNA position under modification is changed according to canonical, Watson–Crick base pairings. The preliminary protocol for the characterization of modified DNA : protein interactions has not yet been applied or tested to verify its operability.

The protocol for the characterization of modified RNA : protein interactions continues to provide valuable insights into the recognition of modified RNAs by various protein readers in ongoing studies. Results derived from this protocol can provide insights

into the role of modifications in biological systems or identify modified RNAs that could be biomarkers for disease. Future improvements by members of the Tamamis lab may include generating robust parametrization for specific modified RNAs and DNAs through FFParam[502] and reducing the possible occurrences of false positives within the protocol.

Study and Design of β-wrapin Binding and Specificity to Amyloidogenic Proteins

The formation of aggregates by amyloidogenic proteins are pathological features of AD, PD, and T2D.[232-239] AD is marked by the presence of senile plaques formed by Aβ[232] and neurofibrillary tangles formed by tau,[233] PD is marked by the presence of Lewy bodies formed by α-syn,[234-238] and T2D is marked by the presence of pancreatic islet amyloid formed by IAPP.[239] In addition to the individual amyloidogenic proteins being implicated in AD, PD, and T2D, increasing evidence also suggest pathological cross talk between these proteins for the three age-related diseases.[251-273] Thus, single-targeted and multi-targeted β-wrapins that sequester monomers of Aβ, α-syn, IAPP, and tau, thereby inhibiting their aggregation could serve as promising potential therapeutic agents for AD, PD, and T2D. Few, subtle changes at the sequence level of β-wrapins greatly change their affinity to the amyloidogenic, and previous attempts to enhance β-wrapins' single-targeted and multi-targeted binding have relied on phage display library experiments and researcher intuition.

To study and design β-wrapin variants for enhanced single-targeted and multi-targeted binding to amyloidogenic proteins, MD simulations and free energy calculations were performed and combined with an in-house developed optimization-based protein

design program that accounts for both single-targeted and multi-targeted binding. Through MD simulations and free energy calculations, carefully tuned for the study of β-wrapins in complex with amyloidogenic proteins, the key commonalities leading to multi-targeted binding properties of specific β-wrapins, potential interaction-based switches diminishing β-wrapin affinity for specific amyloidogenic proteins, and key energetic driving determinants leading to β-wrapin binding and specificity were revealed.[74, 75] From the optimization-based design program, combined with MD simulations, free energy calculations, and structural analysis, promising designed β-wrapin variants with enhanced predicted single-targeted affinity for IAPP, multi-targeted affinity for Aβ, α-syn, and IAPP, and multi-targeted affinity for Aβ and tau were identified. Additionally, rationally designed β-wrapins with N-termini modified with aromatic canonical and non-canonical amino acids that mimic the amyloid inhibition of action of curcumin for Aβ were predicted to encompass both amyloidogenic protein sequestration and amyloid fibril elongation inhibition properties, indicating the potential to amplify the anti-amyloid properties of the β-wrapins.

The MD simulations and free energy calculations developed as part of this doctoral study[74, 75] could be used to model β-wrapins in complex with novel proteins, with sequence similarity to the studied amyloidogenic proteins, in complex with β-wrapins, gain fundamental knowledge on the common underlying basis of amyloidogenic proteins leading to their ability to form β-hairpins and their affinity for β-wrapins, and gain insights into what interactions can be enhanced to potentially improve β-wrapins binding to amyloidogenic proteins.

Simulating and Analyzing the Binding of Compounds onto Montmorillonite Clays

Montmorillonite clays are a promising adsorbent for several, diverse toxic compounds to mitigate their toxicity.[24, 25] Thus, the development of standardized, systematic computational methods to simulate and investigate the mechanisms of diverse compounds binding to clay surfaces and predict their adsorption free energies is essential. Such methods could allow for the evaluation of montmorillonite clays to bind a compound of interest.

To uncover the mechanisms of compounds binding to montmorillonite clays, a systematic computational method to simulate and uncover the mechanisms of diverse compounds binding to clay surfaces and predict their adsorption free energies was developed. The methodology for simulating and uncovering the adsorption of compounds onto montmorillonite clays has been used in several studies investigating the mechanisms of distinct toxic compounds binding to different formulations of montmorillonite clays.[76-78] The systematic generation of simulation files has allowed for the qualitative comparison of compounds' propensity to bind to the clay and the strength of interaction between the bound compounds and the clay, which were used to delineate toxic compounds that can be effectively bound onto the clay from those that cannot.[76-78] Key modes of binding were also identified for the investigated compounds binding to the clay.[76-78] The standardization of MD simulations also allowed for the development of a minimalistic model that predicts the adsorption free energy of individual, diverse compounds for CM.[79]

Future improvements to the methodology of investigating the binding of compounds onto montmorillonite clays by members of the Tamamis Lab may include

expanding its capabilities to study other compounds to clays as well as including additional adsorbent materials beyond clay. The development of adsorbent materials that are effective in sequestering additional compounds is imperative, especially considering that simultaneous subthreshold exposure to various chemicals may result in adverse health effects.[503, 504] Ongoing studies performed in the Tamamis Lab will be used to improve the generalizability of the simulations and analysis programs. The current simulation setup has been modified to allow the study of complex amendments, beyond small compound amendments such as carnitine and choline, to clays and how these complex amendments could enhance the clays' efficacy as an adsorbent.

Elucidating the Pathways of Short Peptide Self- and Co-Assembly

Short self-assembling peptides are gaining increasing interest in materials research as they are easily synthesized or modified, have a wide chemical diversity, and are inherently biocompatible.[22, 23, 429-435, 452] The initial stages of peptide self-assembly and co-assembly is difficult to probe experimentally due to the complex nature of self-assembly and limitations due to sensitivity and time resolution.[80] Difficulties in studying peptide self- and co-assembly has resulted in much of the progress in the design of self- and co-assembly systems being accomplished through trial and error.[80] Gaining information on the early stages of self-assembly can enable the design self-assembled materials with desired functional properties through computational studies.[22, 23, 430-433, 485, 505]

To examine the initial stages of peptide self-assembly and the effect of different solvents, ions, or modifications to the peptides, a suite of structural and energetic analysis

programs tracking and characterizing the formation of clusters within MD simulations were developed. The MD simulations in conjunction with the suite of structural and energetic analysis programs have been successfully implemented in studying the self-assembly of cyclo-HH in different self- and co-assembly environments[81-83] as well as in comparing the self-assembly of different Fmoc-F derivatives.[84] Due to the generalizability of the MD simulations, structural analysis programs, and energy calculation programs, the methods developed may be applied to elucidate the self- and co-assembly of other small molecules, including, but not limited to short peptides, peptide mimetics, nucleic acids, or compounds.

REFERENCES

1. Cleaves, H. J., Molecular Recognition. In *Encyclopedia of Astrobiology*, Gargaud, M.; Amils, R.; Quintanilla, J. C.; Cleaves, H. J.; Irvine, W. M.; Pinti, D. L.; Viso, M., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; pp 1079-1080.

2. Baldridge, K. C.; Contreras, L. M., Functional implications of ribosomal RNA methylation in response to environmental stress. *Crit Rev Biochem Mol Biol* **2014,** *49* (1), 69-89.

3. Morris, K. V.; Mattick, J. S., The rise of regulatory RNA. *Nat Rev Genet* **2014,** *15* (6), 423-37.

4. Gerstberger, S.; Hafner, M.; Tuschl, T., A census of human RNA-binding proteins. *Nat Rev Genet* **2014,** *15* (12), 829-45.

5. Mitchell, S. F.; Parker, R., Principles and properties of eukaryotic mRNPs. *Mol Cell* **2014,** *54* (4), 547-58.

6. Marchese, D.; de Groot, N. S.; Lorenzo Gotor, N.; Livi, C. M.; Tartaglia, G. G., Advances in the characterization of RNA-binding proteins. *Wiley Interdiscip Rev RNA* **2016,** *7* (6), 793-810.

7. Jones, S., Protein-RNA interactions: structural biology and computational modeling techniques. *Biophys Rev* **2016,** *8* (4), 359-367.

8. Mihailovic, M. K.; Chen, A.; Gonzalez-Rivera, J. C.; Contreras, L. M., Defective Ribonucleoproteins, Mistakes in RNA Processing, and Diseases. *Biochemistry* **2017,** *56* (10), 1367-1382.

9. Ji, H. F.; Kong, D. X.; Shen, L.; Chen, L. L.; Ma, B. G.; Zhang, H. Y., Distribution patterns of small-molecule ligands in the protein universe and implications for origin of life and drug discovery. *Genome Biol* **2007,** *8* (8), R176.

10. Whittaker, R. H.; Feeny, P. P., Allelochemics: chemical interactions between species. *Science* **1971,** *171* (3973), 757-70.

11. Dixon, R. A., Natural products and plant disease resistance. *Nature* **2001,** *411* (6839), 843-7.

12. Baldwin, I. T.; Halitschke, R.; Paschold, A.; von Dahl, C. C.; Preston, C. A., Volatile signaling in plant-plant interactions: "talking trees" in the genomics era. *Science* **2006,** *311* (5762), 812-5.

13. Bassler, B. L.; Losick, R., Bacterially speaking. *Cell* **2006,** *125* (2), 237-46.

14. Camilli, A.; Bassler, B. L., Bacterial small-molecule signaling pathways. *Science* **2006,** *311* (5764), 1113-6.

15. Keller, L.; Surette, M. G., Communication in bacteria: an ecological and evolutionary perspective. *Nat Rev Microbiol* **2006,** *4* (4), 249-58.

16. Ladurner, A. G., Rheostat control of gene expression by metabolites. *Mol Cell* **2006,** *24* (1), 1-11.

17. Marreiros, R.; Muller-Schiffmann, A.; Bader, V.; Selvarajah, S.; Dey, D.; Lingappa, V. R.; Korth, C., Viral capsid assembly as a model for protein aggregation diseases: Active processes catalyzed by cellular assembly machines comprising novel drug targets. *Virus Res* **2015,** *207*, 155-64.

18. Aguzzi, A.; O'Connor, T., Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat Rev Drug Discov* **2010,** *9* (3), 237-48.

19. Voss, K. A.; Dorner, J. W.; Cole, R. J., Amelioration of Aflatoxicosis in Rats by Volclay NF-BC, Microfine Bentonite. *J Food Prot* **1993,** *56* (7), 595-598.

20. Viseras, C.; Aguzzi, C.; Cerezo, P.; Lopez-Galindo, A., Uses of clay minerals in semisolid health care and therapeutic products. *Appl Clay Sci* **2007,** *36* (1-3), 37-50.

21. Lindberg, H.; Hard, T.; Lofblom, J.; Stahl, S., A truncated and dimeric format of an Affibody library on bacteria enables FACS-mediated isolation of amyloid-beta aggregation inhibitors with subnanomolar affinity. *Biotechnol J* **2015,** *10* (11), 1707-18.

22. Jonnalagadda, S. V. R.; Gerace, A. J.; Thai, K.; Johnson, J.; Tsimenidis, K.; Jakubowski, J. M.; Shen, C.; Henderson, K. J.; Tamamis, P.; Gkikas, M., Amyloid Peptide Scaffolds Coordinate with Alzheimer's Disease Drugs. *J Phys Chem B* **2020,** *124* (3), 487-503.

23. Jonnalagadda, S. V. R.; Kokotidou, C.; Orr, A. A.; Fotopoulou, E.; Henderson, K. J.; Choi, C. H.; Lim, W. T.; Choi, S. J.; Jeong, H. K.; Mitraki, A.; Tamamis, P., Computational Design of Functional Amyloid Materials with Cesium Binding, Deposition, and Capture Properties. *J Phys Chem B* **2018,** *122* (30), 7555-7568.

24. Phillips, T. D.; Afriyie-Gyawu, E.; Williams, J.; Huebner, H.; Ankrah, N. A.; Ofori-Adjei, D.; Jolly, P.; Johnson, N.; Taylor, J.; Marroquin-Cardona, A.; Xu, L.; Tang, L.; Wang, J. S., Reducing human exposure to aflatoxin through the use of clay: a review. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* **2008,** *25* (2), 134-45.

25. Huebner, H. J.; Herrera, P.; Phillips, T. D., Clay-Based Interventions for the Control of Chemical and Microbial Hazards in Food and Water. In *Preharvest and Postharvest Food Safety*, 2004; pp 389-402.

26. DeLisi, C., Computers in molecular biology: current applications and emerging trends. *Science* **1988,** *240* (4848), 47-52.

27. Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W., Deep learning for molecular design-a review of the state of the art. *Molecular Systems Design & Engineering* **2019,** *4* (4), 828-849.

28. Hospital, A.; Goni, J. R.; Orozco, M.; Gelpi, J. L., Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem* **2015,** *8*, 37-47.

29. Senn, H. M.; Thiel, W., QM/MM methods for biomolecular systems. *Angew Chem Int Ed Engl* **2009,** *48* (7), 1198-229.

30. Field, M. J.; Bash, P. A.; Karplus, M., A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J Comput Chem* **1990,** *11* (6), 700-733.

31. Levitt, M.; Warshel, A., Computer simulation of protein folding. *Nature* **1975,** *253* (5494), 694-8.

32. Nithin, C.; Ghosh, P.; Bujnicki, J. M., Bioinformatics Tools and Benchmarks for Computational Docking and 3D Structure Prediction of RNA-Protein Complexes. *Genes (Basel)* **2018,** *9* (9), 432.

33. Lee, E. H.; Hsin, J.; Sotomayor, M.; Comellas, G.; Schulten, K., Discovery through the computational microscope. *Structure* **2009,** *17* (10), 1295-306.

34. Hollingsworth, S. A.; Dror, R. O., Molecular Dynamics Simulation for All. *Neuron* **2018,** *99* (6), 1129-1143.

35. Lecot, S.; Chevolot, Y.; Phaner-Goutorbe, M.; Yeromonahos, C., Impact of Silane Monolayers on the Adsorption of Streptavidin on Silica and Its Subsequent Interactions with Biotin: Molecular Dynamics and Steered Molecular Dynamics Simulations. *J Phys Chem B* **2020,** *124* (31), 6786-6796.

36. Bagherpoor Helabad, M.; Volkenandt, S.; Imhof, P., Molecular Dynamics Simulations of a Chimeric Androgen Receptor Protein (SPARKI) Confirm the Importance of the Dimerization Domain on DNA Binding Specificity. *Front Mol Biosci* **2020,** *7*, 4.

37. Willemsen, J. A. R.; Myneni, S. C. B.; Bourg, I. C., Molecular Dynamics Simulations of the Adsorption of Phthalate Esters on Smectite Clay Surfaces. *Journal of Physical Chemistry C* **2019,** *123* (22), 13624-13636.

38. Clark, A. J.; Negron, C.; Hauser, K.; Sun, M.; Wang, L.; Abel, R.; Friesner, R. A., Relative Binding Affinity Prediction of Charge-Changing Sequence Mutations with FEP in Protein-Protein Interfaces. *J Mol Biol* **2019,** *431* (7), 1481-1493.

39. Sharma, M.; Anirudh, C. R., Mechanism of mRNA-STAR domain interaction: Molecular dynamics simulations of Mammalian Quaking STAR protein. *Sci Rep* **2017,** *7* (1), 12567.

40. Panel, N.; Sun, Y. J.; Fuentes, E. J.; Simonson, T., A Simple PB/LIE Free Energy Function Accurately Predicts the Peptide Binding Specificity of the Tiam1 PDZ Domain. *Front Mol Biosci* **2017,** *4*, 65.

41. Maffeo, C.; Aksimentiev, A., Molecular mechanism of DNA association with single-stranded DNA binding protein. *Nucleic Acids Res* **2017,** *45* (21), 12125-12139.

42. Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D., Jr., Impact of 2'-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J Comput Chem* **2011,** *32* (9), 1929-43.

43. Darian, E.; Gannett, P. M., Application of molecular dynamics simulations to spin-labeled oligonucleotides. *J Biomol Struct Dyn* **2005,** *22* (5), 579-93.

44. Perez, A.; Morrone, J. A.; Simmerling, C.; Dill, K. A., Advances in free-energy-based simulations of protein folding and ligand binding. *Curr Opin Struct Biol* **2016,** *36*, 25-31.

45. Geng, H.; Chen, F.; Ye, J.; Jiang, F., Applications of Molecular Dynamics Simulation in Structure Prediction of Peptides and Proteins. *Comput Struct Biotechnol J* **2019,** *17*, 1162-1170.

46. Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S., Alchemical free energy methods for drug discovery: progress and challenges. *Curr Opin Struct Biol* **2011,** *21* (2), 150-60.

47. Mobley, D. L.; Dill, K. A., Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure* **2009,** *17* (4), 489-98.

48. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.;

Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc* **2015,** *137* (7), 2695-703.

49. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* **2000,** *33* (12), 889-97.

50. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* **2011,** *51* (1), 69-82.

51. Malde, A. K.; Mark, A. E., Challenges in the determination of the binding modes of non-standard ligands in X-ray crystal complexes. *J Comput Aided Mol Des* **2011,** *25* (1), 1-12.

52. Malde, A. K.; Mark, A. E., 179 Validation of ligands in X-ray crystal structures. *Journal of Biomolecular Structure and Dynamics* **2015,** *33* (sup1), 117-177.

53. Ramirez, D.; Caballero, J., Is It Reliable to Use Common Molecular Docking Methods for Comparing the Binding Affinities of Enantiomer Pairs for Their Protein Target? *Int J Mol Sci* **2016,** *17* (4), 525.

54. Orr, A. A.; Jayaraman, A.; Tamamis, P., Molecular Modeling of Chemoreceptor: Ligand Interactions. In *Bacterial Chemosensing*, Springer: 2018; pp 353-372.

55. Orr, A. A.; Yang, J.; Sule, N.; Chawla, R.; Hull, K. G.; Zhu, M.; Romo, D.; Lele, P. P.; Jayaraman, A.; Manson, M. D.; Tamamis, P., Molecular Mechanism for Attractant Signaling to DHMA by E. coli Tsr. *Biophys J* **2020,** *118* (2), 492-504.

56. Cheng, Y.; Jin, U.-H.; Davidson, L. A.; Chapkin, R. S.; Jayaraman, A.; Tamamis, P.; Orr, A.; Allred, C.; Denison, M. S.; Soshilov, A., Editor's Highlight: Microbial-Derived 1, 4-Dihydroxy-2-naphthoic Acid and Related Compounds as Aryl Hydrocarbon Receptor Agonists/Antagonists: Structure–Activity Relationships and Receptor Modeling. *Toxicological Sciences* **2016,** *155* (2), 458-473.

57. Jin, U. H.; Park, H.; Li, X.; Davidson, L. A.; Allred, C.; Patil, B.; Jayaprakasha, G.; Orr, A. A.; Mao, L.; Chapkin, R. S.; Jayaraman, A.; Tamamis, P.; Safe, S., Structure-Dependent Modulation of Aryl Hydrocarbon Receptor-Mediated Activities by Flavonoids. *Toxicol Sci* **2018,** *164* (1), 205-217.

58. Park, H.; Jin, U. H.; Orr, A. A.; Echegaray, S. P.; Davidson, L. A.; Allred, C. D.; Chapkin, R. S.; Jayaraman, A.; Lee, K.; Tamamis, P.; Safe, S., Isoflavones as Ah Receptor Agonists in Colon-Derived Cell Lines: Structure-Activity Relationships. *Chem Res Toxicol* **2019,** *32* (11), 2353-2364.

59. Yoon, K.; Chen, C. C.; Orr, A. A.; Barreto, P. N.; Tamamis, P.; Safe, S., Activation of COUP-TFI by a Novel Diindolylmethane Derivative. *Cells* **2019,** *8* (3), 220.

60. Boccaletto, P.; Machnicka, M. A.; Purta, E.; Piatkowski, P.; Baginski, B.; Wirecki, T. K.; de Crecy-Lagard, V.; Ross, R.; Limbach, P. A.; Kotter, A.; Helm, M.; Bujnicki, J. M., MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **2018,** *46* (D1), D303-D307.

61. Helm, M.; Motorin, Y., Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat Rev Genet* **2017,** *18* (5), 275-291.

62. Frye, M.; Jaffrey, S. R.; Pan, T.; Rechavi, G.; Suzuki, T., RNA modifications: what have we learned and where are we headed? *Nat Rev Genet* **2016,** *17* (6), 365-72.

63. Pacini, C.; Koziol, M. J., Bioinformatics challenges and perspectives when studying the effect of epigenetic modifications on alternative splicing. *Philos Trans R Soc Lond B Biol Sci* **2018,** *373* (1748).

64. Schaefer, M.; Kapoor, U.; Jantsch, M. F., Understanding RNA modifications: the promises and technological bottlenecks of the 'epitranscriptome'. *Open Biol* **2017,** *7* (5).

65. Orr, A. A.; Gonzalez-Rivera, J. C.; Wilson, M.; Bhikha, P. R.; Wang, D.; Contreras, L. M.; Tamamis, P., A high-throughput and rapid computational method for screening of RNA post-transcriptional modifications that can be recognized by target proteins. *Methods* **2018,** *143*, 34-47.

66. Gonzalez-Rivera, J. C.; Orr, A. A.; Engels, S. M.; Jakubowski, J. M.; Sherman, M. W.; O'Connor, K. N.; Matteson, T.; Woodcock, B. C.; Contreras, L. M.; Tamamis, P., Computational evolution of an RNA-binding protein towards enhanced oxidized-RNA binding. *Comput Struct Biotechnol J* **2020,** *18*, 137-152.

67. Gonzalez-Rivera, J. C.; Orr, A. A.; O'Connor, K. N.; Woodcock, B. C.; Griffith, J.; Tamamis, P.; Contreras, L. M., Broader recognition of RNA modifications among protein readers uncovered using virtual screening.

68. Gruning, C. S. R.; Mirecka, E. A.; Klein, A. N.; Mandelkow, E.; Willbold, D.; Marino, S. F.; Stoldt, M.; Hoyer, W., Alternative conformations of the Tau repeat

domain in complex with an engineered binding protein. *J Biol Chem* **2014,** *289* (33), 23209-23218.

69. Mirecka, E. A.; Feuerstein, S.; Gremer, L.; Schroder, G. F.; Stoldt, M.; Willbold, D.; Hoyer, W., beta-Hairpin of Islet Amyloid Polypeptide Bound to an Aggregation Inhibitor. *Sci Rep* **2016,** *6*, 33474.

70. Mirecka, E. A.; Shaykhalishahi, H.; Gauhar, A.; Akgul, S.; Lecher, J.; Willbold, D.; Stoldt, M.; Hoyer, W., Sequestration of a beta-hairpin for control of alpha-synuclein aggregation. *Angew Chem Int Ed Engl* **2014,** *53* (16), 4227-30.

71. Gronwall, C.; Jonsson, A.; Lindstrom, S.; Gunneriusson, E.; Stahl, S.; Herne, N., Selection and characterization of Affibody ligands binding to Alzheimer amyloid beta peptides. *J Biotechnol* **2007,** *128* (1), 162-83.

72. Hoyer, W.; Gronwall, C.; Jonsson, A.; Stahl, S.; Hard, T., Stabilization of a beta-hairpin in monomeric Alzheimer's amyloid-beta peptide inhibits amyloid formation. *Proc Natl Acad Sci U S A* **2008,** *105* (13), 5099-104.

73. Shaykhalishahi, H.; Mirecka, E. A.; Gauhar, A.; Gruning, C. S.; Willbold, D.; Hard, T.; Stoldt, M.; Hoyer, W., A beta-hairpin-binding protein for three different disease-related amyloidogenic proteins. *Chembiochem* **2015,** *16* (3), 411-4.

74. Orr, A. A.; Shaykhalishahi, H.; Mirecka, E. A.; Jonnalagadda, S. V. R.; Hoyer, W.; Tamamis, P., Elucidating the multi-targeted anti-amyloid activity and enhanced islet amyloid polypeptide binding of beta-wrapins. *Comput Chem Eng* **2018,** *116*, 322-332.

75. Orr, A. A.; Wordehoff, M. M.; Hoyer, W.; Tamamis, P., Uncovering the Binding and Specificity of beta-Wrapins for Amyloid-beta and alpha-Synuclein. *J Phys Chem B* **2016,** *120* (50), 12781-12794.

76. Orr, A. A.; He, S.; Wang, M.; Goodall, A.; Hearon, S. E.; Phillips, T. D.; Tamamis, P., Insights into the interactions of bisphenol and phthalate compounds with unamended and carnitine-amended montmorillonite clays. *Comput Chem Eng* **2020,** *143*, 107063.

77. Wang, M.; Orr, A. A.; He, S.; Dalaijamts, C.; Chiu, W. A.; Tamamis, P.; Phillips, T. D., Montmorillonites Can Tightly Bind Glyphosate and Paraquat Reducing Toxin Exposures and Toxicity. *ACS omega* **2019,** *4* (18), 17702-17713.

78. Wang, M.; Orr, A. A.; Jakubowski, J. M.; Bird, K. E.; Casey, C. M.; Hearon, S. E.; Tamamis, P.; Phillips, T. D., Enhanced adsorption of per- and polyfluoroalkyl substances (PFAS) by edible, nutrient-amended montmorillonite clays. *Water Res* **2021,** *188*, 116534.

79. Orr, A. A.; Wang, M.; Beykal, B.; Ganesh, H. S.; Hearon, S. E.; Pistikopoulos, E. N.; Phillips, T. D.; Tamamis, P., A Combination of Experimental Isotherms and Minimalistic Simulations and Modes for Understanding and Predicting Chemical Adsorption onto Montmorillonite Clays. *ACS Omega* **2021**.

80. Frederix, P.; Patmanidis, I.; Marrink, S. J., Molecular simulations of self-assembling bio-inspired supramolecular systems and their connection to experiments. *Chem Soc Rev* **2018,** *47* (10), 3470-3489.

81. Tao, K.; Chen, Y.; Orr, A. A.; Tian, Z.; Makam, P.; Gilead, S.; Si, M.; Rencus-Lazar, S.; Qu, S.; Zhang, M.; Tamamis, P.; Gazit, E., Enhanced Fluorescence for Bioassembly by Environment-Switching Doping of Metal Ions. *Adv Funct Mater* **2020,** *30* (10).

82. Chen, Y.; Orr, A. A.; Tao, K.; Wang, Z.; Ruggiero, A.; Shimon, L. J. W.; Schnaider, L.; Goodall, A.; Rencus-Lazar, S.; Gilead, S.; Slutsky, I.; Tamamis, P.; Tan, Z.; Gazit, E., High-Efficiency Fluorescence through Bioinspired Supramolecular Self-Assembly. *ACS Nano* **2020,** *14* (3), 2798-2807.

83. Gazit, E.; Chen, Y.; Yang, Y.; Orr, A. A.; Makam, P.; Redko, B.; Haimov, E.; Wang, Y.; Shimon, L. J. W.; Rencus-Lazar, S.; Ju, M.; Tamamis, P.; Dong, H., Self-assembled peptide nano-superstructure towards enzyme mimicking hydrolysis. *Angew Chem Int Ed Engl* **2021,** *n/a* (n/a).

84. Aviv, M.; Cohen-Gerassi, D.; Orr, A. A.; Misra, R.; Shimon, L. J. W.; Tamamis, P.; Adler-Abramovich, L., A Single Fluorine Positioning Affects the Physical Properties of Low Molecular Weight Hydrogels.

85. *Protein-Ligand Interactions: From Molecular Recognition to Drug Design*. 2003.

86. Babine, R. E.; Bender, S. L., Molecular Recognition of Protein−Ligand Complexes: Applications to Drug Design. *Chem Rev* **1997,** *97* (5), 1359-1472.

87. Wallace, A. C.; Laskowski, R. A.; Singh, J.; Thornton, J. M., Molecular recognition by proteins: protein-ligand interactions from a structural perspective. *Biochem Soc Trans* **1996,** *24* (1), 280-4.

88. Xing, F.; Song, G.; Ren, J.; Chaires, J. B.; Qu, X., Molecular recognition of nucleic acids: coralyne binds strongly to poly(A). *FEBS Lett* **2005,** *579* (22), 5035-9.

89. Furukawa, Y.; Suzuki, Y.; Fukuoka, M.; Nagasawa, K.; Nakagome, K.; Shimizu, H.; Mukaiyama, A.; Akiyama, S., A molecular mechanism realizing sequence-specific recognition of nucleic acids by TDP-43. *Sci Rep* **2016,** *6* (1), 20576.

90. Zacco, E.; Martin, S. R.; Thorogate, R.; Pastore, A., The RNA-Recognition Motifs of TAR DNA-Binding Protein 43 May Play a Role in the Aberrant Self-Assembly of the Protein. *Frontiers in Molecular Neuroscience* **2018,** *11* (372).

91. Ball, J. R.; Dimaano, C.; Bilak, A.; Kurchan, E.; Zundel, M. T.; Ullman, K. S., Sequence Preference in RNA Recognition by the Nucleoporin Nup153*. *Journal of Biological Chemistry* **2007,** *282* (12), 8734-8740.

92. Kligun, E.; Mandel-Gutfreund, Y., The role of RNA conformation in RNA-protein recognition. *RNA Biol* **2015,** *12* (7), 720-727.

93. Frejd, F. Y.; Kim, K.-T., Affibody molecules as engineered protein drugs. *Experimental & Molecular Medicine* **2017,** *49* (3), e306-e306.

94. Boutajangout, A.; Lindberg, H.; Awwad, A.; Paul, A.; Baitalmal, R.; Almokyad, I.; Höidén-Guthenberg, I.; Gunneriusson, E.; Frejd, F. Y.; Härd, T.; Löfblom, J.; Ståhl, S.; Wisniewski, T., Affibody-Mediated Sequestration of Amyloid β Demonstrates Preventive Efficacy in a Transgenic Alzheimer's Disease Mouse Model. *Frontiers in Aging Neuroscience* **2019,** *11* (64).

95. Luheshi, L. M.; Hoyer, W.; de Barros, T. P.; van Dijk Hard, I.; Brorsson, A. C.; Macao, B.; Persson, C.; Crowther, D. C.; Lomas, D. A.; Stahl, S.; Dobson, C. M.; Hard, T., Sequestration of the Abeta peptide prevents toxicity and promotes degradation in vivo. *PLoS Biol* **2010,** *8* (3), e1000334.

96. Akyüz, S.; Akyüz, T.; Davies, J. E. D., Adsorption of 2,2′-bipyridyl onto sepiolite, attapulgite and smectite group clay minerals from anatolia: The FT-IR and FT-Raman spectra of surface and intercalated species. *Journal of inclusion phenomena and molecular recognition in chemistry* **1994,** *18* (2), 123-135.

97. Fujimoto, N.; Mori, Y.; Yamagishi, A.; Sato, H., Molecular recognition of star-burst tetranuclear Ru(iii) complexes on a chirally modified clay surface. *Chem Commun (Camb)* **2010,** *46* (30), 5473-5475.

98. Yamagishi, A.; Sato, H., Stereochemistry and Molecular Recognition on the Surface of a Smectite Clay Mineral. *Clays Clay Miner* **2012,** *60* (4), 411-419.

99. Zhu, H. Y.; Yamanaka, S., Molecular recognition by Na-loaded alumina pillared clay. *Journal of the Chemical Society, Faraday Transactions* **1997,** *93* (3), 477-480.

100. Huang, W.; Dedzo, G. K.; Stoyanov, S. R.; Lyubimova, O.; Gusarov, S.; Singh, S.; Lao, H.; Kovalenko, A.; Detellier, C., Molecule–Surface Recognition between Heterocyclic Aromatic Compounds and Kaolinite in Toluene Investigated by Molecular Theory of Solvation and Thermodynamic and Kinetic Experiments. *J. Phys. Chem. C* **2014,** *118* (41), 23821-23834.

101. Haino, T., Molecular-recognition-directed formation of supramolecular polymers. *Polym J* **2013,** *45* (4), 363-383.

102. Harada, A.; Kobayashi, R.; Takashima, Y.; Hashidzume, A.; Yamaguchi, H., Macroscopic self-assembly through molecular recognition. *Nature Chemistry* **2011,** *3* (1), 34-37.

103. Uzun, O.; Sanyal, A.; Jeong, Y.; Rotello, V. M., Molecular recognition induced self-assembly of diblock copolymers: microspheres to vesicles. *Macromol Biosci* **2010,** *10* (5), 481-7.

104. Wang, N.; Hao, H.; Lu, H.; Xu, R., Molecular recognition and self-assembly mechanism of cocrystallization processes. *CrystEngComm* **2017,** *19* (27), 3746-3752.

105. Ciferri, A., Molecular recognition mechanisms directing the self-assembly of biological structures. *Soft Matter* **2020,** *16* (39), 8985-8995.

106. Rebek, J., Introduction to the Molecular Recognition and Self-Assembly Special Feature. *Proceedings of the National Academy of Sciences* **2009,** *106* (26), 10423-10424.

107. Hoeppner, A.; Schmitt, L.; Smits, S. H. J., Proteins and Their Ligands: Their Importance and How to Crystallize Them. In *Advanced Topics on Crystal Growth* [Online] Ferreira, S. O., Ed. IntechOpen: 2013.

108. Slabinski, L.; Jaroszewski, L.; Rodrigues, A. P. C.; Rychlewski, L.; Wilson, I. A.; Lesley, S. A.; Godzik, A., The challenge of protein structure determination--lessons from structural genomics. *Protein science : a publication of the Protein Society* **2007,** *16* (11), 2472-2482.

109. Wlodek, S.; Skillman, A. G.; Nicholls, A., Automated ligand placement and refinement with a combined force field and shape potential. *Acta Crystallographica Section D* **2006,** *62* (7), 741-749.

110. Bartuzi, D.; Kaczor, A. A.; Targowska-Duda, K. M.; Matosiuk, D., Recent Advances and Applications of Molecular Docking to G Protein-Coupled Receptors. *Molecules* **2017,** *22* (2), 340.

111. Liu, Z.; Liu, Y.; Zeng, G.; Shao, B.; Chen, M.; Li, Z.; Jiang, Y.; Liu, Y.; Zhang, Y.; Zhong, H., Application of molecular docking for the degradation of organic pollutants in the environmental remediation: A review. *Chemosphere* **2018,** *203*, 139-150.

112. Pagadala, N. S.; Syed, K.; Tuszynski, J., Software for molecular docking: a review. *Biophys Rev* **2017,** *9* (2), 91-102.

113. Satpathy, R., Application of Molecular Docking Methods on Endocrine Disrupting Chemicals: A Review. *Journal of Applied Biotechnology Reports* **2020,** *7* (2), 74-80.

114. Tao, X.; Huang, Y.; Wang, C.; Chen, F.; Yang, L.; Ling, L.; Che, Z.; Chen, X., Recent developments in molecular docking technology applied in food science: a review. *International Journal of Food Science & Technology* **2020,** *55* (1), 33-45.

115. Yu, D.; Ma, X.; Tu, Y.; Lai, L., Both piston-like and rotational motions are present in bacterial chemoreceptor signaling. *Sci Rep* **2015,** *5* (1), 8640.

116. Hayes, J. M.; Skamnaki, V. T.; Archontis, G.; Lamprakis, C.; Sarrou, J.; Bischler, N.; Skaltsounis, A. L.; Zographos, S. E.; Oikonomakos, N. G., Kinetics, in silico docking, molecular dynamics, and MM-GBSA binding studies on prototype indirubins, KT5720, and staurosporine as phosphorylase kinase ATP-binding site inhibitors: the role of water molecules examined. *Proteins* **2011,** *79* (3), 703-19.

117. *Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery*. IGI Global: 2016.

118. Aminpour, M.; Montemagno, C.; Tuszynski, J. A., An Overview of Molecular Modeling for Drug Discovery with Specific Illustrative Examples of Applications. *Molecules* **2019,** *24* (9).

119. Tutumlu, G.; Dogan, B.; Avsar, T.; Orhan, M. D.; Calis, S.; Durdagi, S., Integrating Ligand and Target-Driven Based Virtual Screening Approaches With in vitro Human Cell Line Models and Time-Resolved Fluorescence Resonance Energy Transfer Assay to Identify Novel Hit Compounds Against BCL-2. *Front Chem* **2020,** *8*, 167.

120. Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S., The PDBbind database: methodologies and updates. *J Med Chem* **2005,** *48* (12), 4111-9.

121. Andrusier, N.; Nussinov, R.; Wolfson, H. J., FireDock: fast interaction refinement in molecular docking. *Proteins* **2007,** *69* (1), 139-59.

122. Rastelli, G.; Pinzi, L., Refinement and Rescoring of Virtual Screening Results. *Front Chem* **2019,** *7*, 498.

123. Rastelli, G.; Degliesposti, G.; Del Rio, A.; Sgobba, M., Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening. *Chem Biol Drug Des* **2009,** *73* (3), 283-6.

124. Masetti, M.; Falchi, F.; Gioia, D.; Recanatini, M.; Ciurli, S.; Musiani, F., Targeting the Protein Tunnels of the Urease Accessory Complex: A Theoretical Investigation. *Molecules* **2020,** *25* (12).

125. Pinzi, L.; Lherbet, C.; Baltas, M.; Pellati, F.; Rastelli, G., In Silico Repositioning of Cannabigerol as a Novel Inhibitor of the Enoyl Acyl Carrier Protein (ACP) Reductase (InhA). *Molecules* **2019,** *24* (14).

126. Mamonov, A. B.; Moghadasi, M.; Mirzaei, H.; Zarbafian, S.; Grove, L. E.; Bohnuud, T.; Vakili, P.; Ch. Paschalidis, I.; Vajda, S.; Kozakov, D., Focused grid-based resampling for protein docking and mapping. *J Comput Chem* **2016,** *37* (11), 961-970.

127. Devaurs, D.; Antunes, D. A.; Hall-Swan, S.; Mitchell, N.; Moll, M.; Lizée, G.; Kavraki, L. E., Using parallelized incremental meta-docking can solve the conformational sampling issue when docking large ligands to proteins. *BMC Molecular and Cell Biology* **2019,** *20* (1), 42.

128. Tamamis, P.; Floudas, C. A., Molecular recognition of CXCR4 by a dual tropic HIV-1 gp120 V3 loop. *Biophysical journal* **2013,** *105* (6), 1502-1514.

129. Tamamis, P.; Floudas, C. A., Molecular Recognition of CCR5 by an HIV-1 gp120 V3 Loop. *PLoS ONE* **2014,** *9* (4), e95767.

130. Tamamis, P.; Floudas, C. A., Elucidating a Key Anti-HIV-1 and Cancer-Associated Axis: The Structure of CCL5 (Rantes) in Complex with CCR5. *Sci Rep* **2014,** *4* (1), 5447.

131. Tamamis, P.; Kieslich, C. A.; Nikiforovich, G. V.; Woodruff, T. M.; Morikis, D.; Archontis, G., Insights into the mechanism of C5aR inhibition by PMX53 via implicit solvent molecular dynamics simulations and docking. *BMC Biophys* **2014,** *7*, 5-5.

132. Orr, A. Computational Insights Into Serine and DHMA Binding to the Tsr Chemoreceptor of Escherichia Coli. 2015.

133. Im, W.; Lee, M. S.; Brooks, C. L., 3rd, Generalized born model with a simple smoothing function. *J Comput Chem* **2003,** *24* (14), 1691-702.

134. Giani Tagliabue, S.; Faber, S. C.; Motta, S.; Denison, M. S.; Bonati, L., Modeling the binding of diverse ligands within the Ah receptor ligand binding domain. *Sci Rep* **2019,** *9* (1), 10693.

135. Xing, Y.; Nukaya, M.; Satyshur, K. A.; Jiang, L.; Stanevich, V.; Korkmaz, E. N.; Burdette, L.; Kennedy, G. D.; Cui, Q.; Bradfield, C. A., Identification of the Ah-receptor structural determinants for ligand preferences. *Toxicol Sci* **2012,** *129* (1), 86-97.

136. Bisson, W. H.; Koch, D. C.; O'Donnell, E. F.; Khalil, S. M.; Kerkvliet, N. I.; Tanguay, R. L.; Abagyan, R.; Kolluri, S. K., Modeling of the aryl hydrocarbon receptor

(AhR) ligand binding domain and its utility in virtual ligand screening to predict new AhR ligands. *J Med Chem* **2009,** *52* (18), 5635-41.

137. Jin, U. H.; Cheng, Y.; Park, H.; Davidson, L. A.; Callaway, E. S.; Chapkin, R. S.; Jayaraman, A.; Asante, A.; Allred, C.; Weaver, E. A.; Safe, S., Short Chain Fatty Acids Enhance Aryl Hydrocarbon (Ah) Responsiveness in Mouse Colonocytes and Caco-2 Human Colon Cancer Cells. *Sci Rep* **2017,** *7* (1), 10163.

138. Motto, I.; Bordogna, A.; Soshilov, A. A.; Denison, M. S.; Bonati, L., New aryl hydrocarbon receptor homology model targeted to improve docking reliability. *J Chem Inf Model* **2011,** *51* (11), 2868-81.

139. Pandini, A.; Soshilov, A. A.; Song, Y.; Zhao, J.; Bonati, L.; Denison, M. S., Detection of the TCDD binding-fingerprint within the Ah receptor ligand binding domain by structurally driven mutagenesis and functional analysis. *Biochemistry* **2009,** *48* (25), 5972-83.

140. Samanta, D.; Borbat, P. P.; Dzikovski, B.; Freed, J. H.; Crane, B. R., Bacterial chemoreceptor dynamics correlate with activity state and are coupled over long distances. *Proc Natl Acad Sci U S A* **2015,** *112* (8), 2455-60.

141. Kitanovic, S.; Ames, P.; Parkinson, J. S., Mutational analysis of the control cable that mediates transmembrane signaling in the Escherichia coli serine chemoreceptor. *J Bacteriol* **2011,** *193* (19), 5062-72.

142. Tajima, H.; Imada, K.; Sakuma, M.; Hattori, F.; Nara, T.; Kamo, N.; Homma, M.; Kawagishi, I., Ligand specificity determined by differentially arranged common ligand-binding residues in bacterial amino acid chemoreceptors Tsr and Tar. *J Biol Chem* **2011,** *286* (49), 42200-42210.

143. Hedblom, M. L.; Adler, J., Genetic and biochemical properties of Escherichia coli mutants with defects in serine chemotaxis. *J Bacteriol* **1980,** *144* (3), 1048-1060.

144. Iwasaki, Y. W.; Siomi, M. C.; Siomi, H., PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem* **2015,** *84*, 405-33.

145. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P., From DNA to RNA. In *Molecular Biology of the Cell*, 4th ed.; Garland Science: New York, 2002.

146. Jankowsky, E.; Harris, M. E., Specificity and nonspecificity in RNA-protein interactions. *Nat Rev Mol Cell Biol* **2015,** *16* (9), 533-44.

147. Setzer, D. R., Measuring equilibrium and kinetic constants using gel retardation assays. *Methods Mol Biol* **1999,** *118*, 115-28.

148. Wilson, G. M.; Sutphen, K.; Chuang, K.; Brewer, G., Folding of A+U-rich RNA elements modulates AUF1 binding. Potential roles in regulation of mRNA turnover. *J Biol Chem* **2001,** *276* (12), 8695-704.

149. Walter, N. G.; Hampel, K. J.; Brown, K. M.; Burke, J. M., Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer. *EMBO J* **1998,** *17* (8), 2378-91.

150. Katsamba, P. S.; Park, S.; Laird-Offringa, I. A., Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods* **2002,** *26* (2), 95-104.

151. Martin, D.; Charpilienne, A.; Parent, A.; Boussac, A.; D'Autreaux, B.; Poupon, J.; Poncet, D., The rotavirus nonstructural protein NSP5 coordinates a [2Fe-2S] iron-sulfur cluster that modulates interaction to RNA. *FASEB J* **2013,** *27* (3), 1074-83.

152. Schulz, S.; Doller, A.; Pendini, N. R.; Wilce, J. A.; Pfeilschifter, J.; Eberhardt, W., Domain-specific phosphomimetic mutation allows dissection of different protein kinase C (PKC) isotype-triggered activities of the RNA binding protein HuR. *Cell Signal* **2013,** *25* (12), 2485-95.

153. Sunwoo, H.; Wu, J. Y.; Lee, J. T., The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. *Proc Natl Acad Sci U S A* **2015,** *112* (31), E4216-25.

154. Ascano, M.; Hafner, M.; Cekan, P.; Gerstberger, S.; Tuschl, T., Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip Rev RNA* **2012,** *3* (2), 159-77.

155. Konig, J.; Zarnack, K.; Rot, G.; Curk, T.; Kayikci, M.; Zupan, B.; Turner, D. J.; Luscombe, N. M.; Ule, J., iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp* **2011,** (50).

156. Mann, M., Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* **2006,** *7* (12), 952-8.

157. Jones, S.; Daley, D. T.; Luscombe, N. M.; Berman, H. M.; Thornton, J. M., Protein-RNA interactions: a structural analysis. *Nucleic Acids Res* **2001,** *29* (4), 943-54.

158. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000,** *28* (1), 235-42.

159. Parois, P.; Cooper, R. I.; Thompson, A. L., Crystal structures of increasingly large molecules: meeting the challenges with CRYSTALS software. *Chem Cent J* **2015,** *9*, 30.

160. He, J.; Tao, H.; Huang, S. Y., Protein-ensemble-RNA docking by efficient consideration of protein flexibility through homology models. *Bioinformatics* **2019,** *35* (23), 4994-5002.

161. Kawabata, T., HOMCOS: an updated server to search and model complex 3D structures. *J Struct Funct Genomics* **2016,** *17* (4), 83-99.

162. Li, J.; Feng, Z.; Wu, J.; Huang, Y.; Lu, G.; Zhu, M.; Wang, B.; Mao, X.; Tao, X., Structure and function analysis of nucleocapsid protein of tomato spotted wilt virus interacting with RNA using homology modeling. *J Biol Chem* **2015,** *290* (7), 3950-61.

163. Zhao, H.; Yang, Y.; Janga, S. C.; Kao, C. C.; Zhou, Y., Prediction and validation of the unexplored RNA-binding protein atlas of the human proteome. *Proteins* **2014,** *82* (4), 640-7.

164. Zhao, H.; Yang, Y.; Zhou, Y., Structure-based prediction of RNA-binding domains and RNA-binding sites and application to structural genomics targets. *Nucleic Acids Res* **2011,** *39* (8), 3017-25.

165. Arnautova, Y. A.; Abagyan, R.; Totrov, M., Protein-RNA Docking Using ICM. *J Chem Theory Comput* **2018,** *14* (9), 4971-4984.

166. Delgado Blanco, J.; Radusky, L. G.; Cianferoni, D.; Serrano, L., Protein-assisted RNA fragment docking (RnaX) for modeling RNA-protein interactions using ModelX. *Proc Natl Acad Sci U S A* **2019,** *116* (49), 24568-24573.

167. Tuszynska, I.; Magnus, M.; Jonak, K.; Dawson, W.; Bujnicki, J. M., NPDock: a web server for protein-nucleic acid docking. *Nucleic Acids Res* **2015,** *43* (W1), W425-30.

168. Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S. Y., HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res* **2017,** *45* (W1), W365-W373.

169. Iwakiri, J.; Hamada, M.; Asai, K.; Kameda, T., Improved Accuracy in RNA-Protein Rigid Body Docking by Incorporating Force Field for Molecular Dynamics Simulation into the Scoring Function. *J Chem Theory Comput* **2016,** *12* (9), 4688-97.

170. Chang, S.; Zhang, D. W.; Xu, L.; Wan, H.; Hou, T. J.; Kong, R., Exploring the molecular basis of RNA recognition by the dimeric RNA-binding protein via molecular simulation methods. *RNA Biol* **2016,** *13* (11), 1133-1143.

171. Krepl, M.; Clery, A.; Blatter, M.; Allain, F. H.; Sponer, J., Synergy between NMR measurements and MD simulations of protein/RNA complexes: application to the

RRMs, the most common RNA recognition motifs. *Nucleic Acids Res* **2016,** *44* (13), 6452-70.

172. Krepl, M.; Havrila, M.; Stadlbauer, P.; Banas, P.; Otyepka, M.; Pasulka, J.; Stefl, R.; Sponer, J., Can We Execute Stable Microsecond-Scale Atomistic Simulations of Protein-RNA Complexes? *J Chem Theory Comput* **2015,** *11* (3), 1220-43.

173. Spears, J. L.; Xiao, X.; Hall, C. K.; Agris, P. F., Amino acid signature enables proteins to recognize modified tRNA. *Biochemistry* **2014,** *53* (7), 1125-33.

174. Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P.; Walter, N. G.; Otyepka, M., RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem Rev* **2018,** *118* (8), 4177-4338.

175. Whitford, P. C.; Sanbonmatsu, K. Y., Simulating movement of tRNA through the ribosome during hybrid-state formation. *J Chem Phys* **2013,** *139* (12), 121919.

176. Xiao, X.; Agris, P. F.; Hall, C. K., Designing peptide sequences in flexible chain conformations to bind RNA: a search algorithm combining Monte Carlo, self-consistent mean field and concerted rotation techniques. *J Chem Theory Comput* **2015,** *11* (2), 740-52.

177. Xu, Y.; Vanommeslaeghe, K.; Aleksandrov, A.; MacKerell, A. D., Jr.; Nilsson, L., Additive CHARMM force field for naturally occurring modified ribonucleotides. *J Comput Chem* **2016,** *37* (10), 896-912.

178. Khoury, G. A.; Thompson, J. P.; Smadbeck, J.; Kieslich, C. A.; Floudas, C. A., Forcefield_PTM: Ab Initio Charge and AMBER Forcefield Parameters for Frequently Occurring Post-Translational Modifications. *J Chem Theory Comput* **2013,** *9* (12), 5653-5674.

179. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M., Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* **2007,** *92* (11), 3817-29.

180. Hayes, R. L.; Noel, J. K.; Whitford, P. C.; Mohanty, U.; Sanbonmatsu, K. Y.; Onuchic, J. N., Reduced model captures Mg(2+)-RNA interaction free energy of riboswitches. *Biophys J* **2014,** *106* (7), 1508-19.

181. Roy, S.; Hennelly, S. P.; Lammert, H.; Onuchic, J. N.; Sanbonmatsu, K. Y., Magnesium controls aptamer-expression platform switching in the SAM-I riboswitch. *Nucleic Acids Res* **2019,** *47* (6), 3158-3170.

182. Roy, S.; Onuchic, J. N.; Sanbonmatsu, K. Y., Cooperation between Magnesium and Metabolite Controls Collapse of the SAM-I Riboswitch. *Biophys J* **2017,** *113* (2), 348-359.

183. Kognole, A. A.; MacKerell, A. D., Jr., Contributions and competition of Mg(2+) and K(+) in folding and stabilization of the Twister ribozyme. *RNA* **2020,** *26* (11), 1704-1715.

184. Cook, K. B.; Hughes, T. R.; Morris, Q. D., High-throughput characterization of protein-RNA interactions. *Brief Funct Genomics* **2015,** *14* (1), 74-89.

185. Klus, P.; Bolognesi, B.; Agostini, F.; Marchese, D.; Zanzoni, A.; Tartaglia, G. G., The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities. *Bioinformatics* **2014,** *30* (11), 1601-8.

186. Livi, C. M.; Blanzieri, E., Protein-specific prediction of mRNA binding using RNA sequences, binding motifs and predicted secondary structures. *BMC Bioinformatics* **2014,** *15*, 123.

187. Xiao, X.; Agris, P. F.; Hall, C. K., Introducing folding stability into the score function for computational design of RNA-binding peptides boosts the probability of success. *Proteins* **2016,** *84* (5), 700-11.

188. Xiao, X.; Hall, C. K.; Agris, P. F., The design of a peptide sequence to inhibit HIV replication: a search algorithm combining Monte Carlo and self-consistent mean field techniques. *J Biomol Struct Dyn* **2014,** *32* (10), 1523-36.

189. Xiao, X.; Hung, M. E.; Leonard, J. N.; Hall, C. K., Adding energy minimization strategy to peptide-design algorithm enables better search for RNA-binding peptides: Redesigned lambda N peptide binds boxB RNA. *J Comput Chem* **2016,** *37* (27), 2423-35.

190. Lewis, C. J.; Pan, T.; Kalsotra, A., RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat Rev Mol Cell Biol* **2017,** *18* (3), 202-210.

191. Frohlich, K. M.; Sarachan, K. L.; Todd, G. C.; Basanta-Sanchez, M.; Väre, V. Y. P.; Agris, P. F., Post-Transcriptional Modifications of RNA: Impact on RNA Function and Human Health. In *Modified nucleic acids in biology and medicine*, Jurga, S.; Erdmann, V. A.; Barciszewski, J., Eds. Springer International Publishing: Cham, 2016; pp 91-130.

192. Khoury, G. A.; Smadbeck, J.; Tamamis, P.; Vandris, A. C.; Kieslich, C. A.; Floudas, C. A., Forcefield_NCAA: ab initio charge parameters to aid in the discovery and design of therapeutic proteins and peptides with unnatural amino acids and their

application to complement inhibitors of the compstatin family. *ACS Synth Biol* **2014,** *3* (12), 855-69.

193. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **2010,** *31* (4), 671-90.

194. Vanommeslaeghe, K.; MacKerell, A. D., Jr., Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J Chem Inf Model* **2012,** *52* (12), 3144-54.

195. Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr., Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J Chem Inf Model* **2012,** *52* (12), 3155-68.

196. Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D., Jr., Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J Comput Chem* **2012,** *33* (31), 2451-68.

197. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: the biomolecular simulation program. *J Comput Chem* **2009,** *30* (10), 1545-614.

198. Huang, J.; MacKerell, A. D., Jr., CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **2013,** *34* (25), 2135-45.

199. Aduri, R.; Psciuk, B. T.; Saro, P.; Taniga, H.; Schlegel, H. B.; SantaLucia, J., AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. *J Chem Theory Comput* **2007,** *3* (4), 1464-75.

200. Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* **2020,** *16* (1), 528-552.

201. Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr., Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009,** *77* (4), 778-95.

202. Tamamis, P.; Morikis, D.; Floudas, C. A.; Archontis, G., Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynamics simulations. *Proteins* **2010,** *78* (12), 2655-67.

203. Feig, M.; Mirjalili, V., Protein structure refinement via molecular-dynamics simulations: What works and what does not? *Proteins* **2016,** *84 Suppl 1* (S1), 282-92.

204. Feig, M., Local Protein Structure Refinement via Molecular Dynamics Simulations with locPREFMD. *J Chem Inf Model* **2016,** *56* (7), 1304-12.

205. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y., The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **2015,** *12* (1), 7-8.

206. Webb, B.; Sali, A., Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **2016,** *54*, 5 6 1-5 6 37.

207. Boniecki, M. J.; Lach, G.; Dawson, W. K.; Tomala, K.; Lukasz, P.; Soltysinski, T.; Rother, K. M.; Bujnicki, J. M., SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res* **2016,** *44* (7), e63.

208. Reuter, J. S.; Mathews, D. H., RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **2010,** *11* (1), 129.

209. Xu, X.; Zhao, P.; Chen, S. J., Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS ONE* **2014,** *9* (9), e107504.

210. Lee, K. H.; Chen, J., Optimization of the GBMV2 implicit solvent force field for accurate simulation of protein conformational equilibria. *J Comput Chem* **2017,** *38* (16), 1332-1341.

211. Stickney, L. M.; Hankins, J. S.; Miao, X.; Mackie, G. A., Function of the conserved S1 and KH domains in polynucleotide phosphorylase. *J Bacteriol* **2005,** *187* (21), 7214-21.

212. Symmons, M. F.; Jones, G. H.; Luisi, B. F., A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure* **2000,** *8* (11), 1215-1226.

213. Wu, J.; Jiang, Z.; Liu, M.; Gong, X.; Wu, S.; Burns, C. M.; Li, Z., Polynucleotide phosphorylase protects Escherichia coli against oxidative stress. *Biochemistry* **2009,** *48* (9), 2012-20.

214. Zhang, Y.; Qian, D.; Li, Z.; Huang, Y.; Wu, Q.; Ru, G.; Chen, M.; Wang, B., Oxidative stress-induced DNA damage of mouse zygotes triggers G2/M checkpoint and phosphorylates Cdc25 and Cdc2. *Cell Stress Chaperones* **2016,** *21* (4), 687-96.

215. Dressaire, C.; Pobre, V.; Laguerre, S.; Girbal, L.; Arraiano, C. M.; Cocaign-Bousquet, M., PNPase is involved in the coordination of mRNA degradation and expression in stationary phase cells of Escherichia coli. *BMC Genomics* **2018,** *19* (1), 848.

216. Nurmohamed, S.; Vaidialingam, B.; Callaghan, A. J.; Luisi, B. F., Crystal structure of Escherichia coli polynucleotide phosphorylase core bound to RNase E, RNA and manganese: implications for catalytic mechanism and RNA degradosome assembly. *J Mol Biol* **2009,** *389* (1), 17-33.

217. Hardwick, S. W.; Gubbey, T.; Hug, I.; Jenal, U.; Luisi, B. F., Crystal structure of Caulobacter crescentus polynucleotide phosphorylase reveals a mechanism of RNA substrate channelling and RNA degradosome assembly. *Open Biol* **2012,** *2* (4), 120028.

218. UniProt, C., UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **2019,** *47* (D1), D506-D515.

219. Smadbeck, J.; Peterson, M. B.; Khoury, G. A.; Taylor, M. S.; Floudas, C. A., Protein WISDOM: a workbench for in silico de novo design of biomolecules. *J Vis Exp* **2013,** (77).

220. Lukavsky, P. J.; Daujotyte, D.; Tollervey, J. R.; Ule, J.; Stuani, C.; Buratti, E.; Baralle, F. E.; Damberger, F. F.; Allain, F. H., Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat Struct Mol Biol* **2013,** *20* (12), 1443-9.

221. Teplova, M.; Malinina, L.; Darnell, J. C.; Song, J.; Lu, M.; Abagyan, R.; Musunuru, K.; Teplov, A.; Burley, S. K.; Darnell, R. B.; Patel, D. J., Protein-RNA and protein-protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. *Structure* **2011,** *19* (7), 930-44.

222. Xu, C.; Liu, K.; Ahmed, H.; Loppnau, P.; Schapira, M.; Min, J., Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. *J Biol Chem* **2015,** *290* (41), 24902-13.

223. Arai, T.; Hasegawa, M.; Akiyama, H.; Ikeda, K.; Nonaka, T.; Mori, H.; Mann, D.; Tsuchiya, K.; Yoshida, M.; Hashizume, Y.; Oda, T., TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem Biophys Res Commun* **2006,** *351* (3), 602-11.

224. Bai, Y.; Yang, C.; Wu, R.; Huang, L.; Song, S.; Li, W.; Yan, P.; Lin, C.; Li, D.; Zhang, Y., YTHDF1 Regulates Tumorigenicity and Cancer Stem Cell-Like Activity in Human Colorectal Carcinoma. *Front Oncol* **2019,** *9*, 332.

225. Bose, J. K.; Wang, I. F.; Hung, L.; Tarn, W. Y.; Shen, C. K., TDP-43 overexpression enhances exon 7 inclusion during the survival of motor neuron pre-mRNA splicing. *J Biol Chem* **2008,** *283* (43), 28852-9.

226. Buratti, E.; Baralle, F. E., Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J Biol Chem* **2001,** *276* (39), 36337-43.

227. Eaton, A.; Bernier, F. P.; Goedhart, C.; Caluseriu, O.; Lamont, R. E.; Boycott, K. M.; Parboosingh, J. S.; Innes, A. M.; Care4Rare Canada, C., Is PNPT1-related hearing loss ever non-syndromic? Whole exome sequencing of adult siblings expands the natural history of PNPT1-related disorders. *Am J Med Genet A* **2018,** *176* (11), 2487-2493.

228. Lu, W.; Tirumuru, N.; St Gelais, C.; Koneru, P. C.; Liu, C.; Kvaratskhelia, M.; He, C.; Wu, L., N(6)-Methyladenosine-binding proteins suppress HIV-1 infectivity and viral production. *J Biol Chem* **2018,** *293* (34), 12992-13005.

229. Matilainen, S.; Carroll, C. J.; Richter, U.; Euro, L.; Pohjanpelto, M.; Paetau, A.; Isohanni, P.; Suomalainen, A., Defective mitochondrial RNA processing due to PNPT1 variants causes Leigh syndrome. *Hum Mol Genet* **2017,** *26* (17), 3352-3361.

230. Neumann, M.; Sampathu, D. M.; Kwong, L. K.; Truax, A. C.; Micsenyi, M. C.; Chou, T. T.; Bruce, J.; Schuck, T.; Grossman, M.; Clark, C. M.; McCluskey, L. F.; Miller, B. L.; Masliah, E.; Mackenzie, I. R.; Feldman, H.; Feiden, W.; Kretzschmar, H. A.; Trojanowski, J. Q.; Lee, V. M., Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **2006,** *314* (5796), 130-3.

231. Xin, Y.; Li, Z.; Zheng, H.; Ho, J.; Chan, M. T. V.; Wu, W. K. K., Neuro-oncological ventral antigen 1 (NOVA1): Implications in neurological diseases and cancers. *Cell Prolif* **2017,** *50* (4).

232. Glenner, G. G.; Wong, C. W., Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem Biophys Res Commun* **1984,** *120* (3), 885-90.

233. Grundke-Iqbal, I.; Iqbal, K.; Quinlan, M.; Tung, Y. C.; Zaidi, M. S.; Wisniewski, H. M., Microtubule-associated protein tau. A component of Alzheimer paired helical filaments. *J Biol Chem* **1986,** *261* (13), 6084-9.

234. Braak, H.; Del Tredici, K.; Rub, U.; de Vos, R. A.; Jansen Steur, E. N.; Braak, E., Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging* **2003,** *24* (2), 197-211.

235. Del Tredici, K.; Rub, U.; De Vos, R. A.; Bohl, J. R.; Braak, H., Where does parkinson disease pathology begin in the brain? *J Neuropathol Exp Neurol* **2002,** *61* (5), 413-26.

236. Polymeropoulos, M. H.; Lavedan, C.; Leroy, E.; Ide, S. E.; Dehejia, A.; Dutra, A.; Pike, B.; Root, H.; Rubenstein, J.; Boyer, R.; Stenroos, E. S.; Chandrasekharappa, S.; Athanassiadou, A.; Papapetropoulos, T.; Johnson, W. G.; Lazzarini, A. M.; Duvoisin, R. C.; Di Iorio, G.; Golbe, L. I.; Nussbaum, R. L., Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **1997,** *276* (5321), 2045-7.

237. Spillantini, M. G.; Schmidt, M. L.; Lee, V. M.; Trojanowski, J. Q.; Jakes, R.; Goedert, M., Alpha-synuclein in Lewy bodies. *Nature* **1997,** *388* (6645), 839-40.

238. Kraybill, M. L.; Larson, E. B.; Tsuang, D. W.; Teri, L.; McCormick, W. C.; Bowen, J. D.; Kukull, W. A.; Leverenz, J. B.; Cherrier, M. M., Cognitive differences in dementia patients with autopsy-verified AD, Lewy body pathology, or both. *Neurology* **2005,** *64* (12), 2069-73.

239. Ehrlich, J. C.; Ratner, I. M., Amyloidosis of the islets of Langerhans. A restudy of islet hyalin in diabetic and non-diabetic individuals. *Am J Pathol* **1961,** *38*, 49-59.

240. Falcon, B.; Zhang, W.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Vidal, R.; Crowther, R. A.; Ghetti, B.; Scheres, S. H. W.; Goedert, M., Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* **2018,** *561* (7721), 137-140.

241. Falcon, B.; Zhang, W.; Schweighauser, M.; Murzin, A. G.; Vidal, R.; Garringer, H. J.; Ghetti, B.; Scheres, S. H. W.; Goedert, M., Tau filaments from multiple cases of sporadic and inherited Alzheimer's disease adopt a common fold. *Acta Neuropathol* **2018,** *136* (5), 699-708.

242. Falcon, B.; Zivanov, J.; Zhang, W.; Murzin, A. G.; Garringer, H. J.; Vidal, R.; Crowther, R. A.; Newell, K. L.; Ghetti, B.; Goedert, M.; Scheres, S. H. W., Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **2019,** *568* (7752), 420-423.

243. Fitzpatrick, A. W. P.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H. W., Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017,** *547* (7662), 185-190.

244. Gremer, L.; Scholzel, D.; Schenk, C.; Reinartz, E.; Labahn, J.; Ravelli, R. B. G.; Tusche, M.; Lopez-Iglesias, C.; Hoyer, W.; Heise, H.; Willbold, D.; Schroder, G. F., Fibril structure of amyloid-beta(1-42) by cryo-electron microscopy. *Science* **2017,** *358* (6359), 116-119.

245. Guerrero-Ferreira, R.; Taylor, N. M.; Mona, D.; Ringler, P.; Lauer, M. E.; Riek, R.; Britschgi, M.; Stahlberg, H., Cryo-EM structure of alpha-synuclein fibrils. *elife* **2018,** *7*.

246. Li, B.; Ge, P.; Murray, K. A.; Sheth, P.; Zhang, M.; Nair, G.; Sawaya, M. R.; Shin, W. S.; Boyer, D. R.; Ye, S.; Eisenberg, D. S.; Zhou, Z. H.; Jiang, L., Cryo-EM of full-length alpha-synuclein reveals fibril polymorphs with a common structural kernel. *Nat Commun* **2018,** *9* (1), 3609.

247. Li, Y.; Zhao, C.; Luo, F.; Liu, Z.; Gui, X.; Luo, Z.; Zhang, X.; Li, D.; Liu, C.; Li, X., Amyloid fibril structure of alpha-synuclein determined by cryo-electron microscopy. *Cell Res* **2018,** *28* (9), 897-903.

248. Ni, X.; McGlinchey, R. P.; Jiang, J.; Lee, J. C., Structural Insights into alpha-Synuclein Fibril Polymorphism: Effects of Parkinson's Disease-Related C-Terminal Truncations. *J Mol Biol* **2019,** *431* (19), 3913-3919.

249. Roder, C.; Kupreichyk, T.; Gremer, L.; Schafer, L. U.; Pothula, K. R.; Ravelli, R. B. G.; Willbold, D.; Hoyer, W.; Schroder, G. F., Cryo-EM structure of islet amyloid polypeptide fibrils reveals similarities with amyloid-beta fibrils. *Nat Struct Mol Biol* **2020,** *27* (7), 660-667.

250. Hard, T.; Lendel, C., Inhibition of amyloid formation. *J Mol Biol* **2012,** *421* (4-5), 441-65.

251. Clinton, L. K.; Blurton-Jones, M.; Myczek, K.; Trojanowski, J. Q.; LaFerla, F. M., Synergistic Interactions between Abeta, tau, and alpha-synuclein: acceleration of neuropathology and cognitive decline. *J Neurosci* **2010,** *30* (21), 7281-9.

252. Hamilton, R. L., Lewy bodies in Alzheimer's disease: a neuropathological review of 145 cases using alpha-synuclein immunohistochemistry. *Brain Pathol* **2000,** *10* (3), 378-84.

253. Akter, K.; Lanza, E. A.; Martin, S. A.; Myronyuk, N.; Rua, M.; Raffa, R. B., Diabetes mellitus and Alzheimer's disease: shared pathology and treatment? *Br J Clin Pharmacol* **2011,** *71* (3), 365-76.

254. Barbagallo, M.; Dominguez, L. J., Type 2 diabetes mellitus and Alzheimer's disease. *World J Diabetes* **2014,** *5* (6), 889-93.

255. Crane, P. K.; Walker, R.; Hubbard, R. A.; Li, G.; Nathan, D. M.; Zheng, H.; Haneuse, S.; Craft, S.; Montine, T. J.; Kahn, S. E.; McCormick, W.; McCurry, S. M.; Bowen, J. D.; Larson, E. B., Glucose levels and risk of dementia. *N Engl J Med* **2013,** *369* (6), 540-8.

256. Janson, J.; Laedtke, T.; Parisi, J. E.; O'Brien, P.; Petersen, R. C.; Butler, P. C., Increased risk of type 2 diabetes in Alzheimer disease. *Diabetes* **2004,** *53* (2), 474-81.

257. Kroner, Z., The relationship between Alzheimer's disease and diabetes: Type 3 diabetes? *Altern Med Rev* **2009,** *14* (4), 373-379.

258. Moreira, I. S.; Koukos, P. I.; Melo, R.; Almeida, J. G.; Preto, A. J.; Schaarschmidt, J.; Trellet, M.; Gumus, Z. H.; Costa, J.; Bonvin, A., SpotOn: High Accuracy Identification of Protein-Protein Interface Hot-Spots. *Sci Rep* **2017,** *7* (1), 8007.

259. Oskarsson, M. E.; Paulsson, J. F.; Schultz, S. W.; Ingelsson, M.; Westermark, P.; Westermark, G. T., In vivo seeding and cross-seeding of localized amyloidosis: a molecular link between type 2 diabetes and Alzheimer disease. *Am J Pathol* **2015,** *185* (3), 834-46.

260. Sridhar, G. R.; Lakshmi, G.; Nagamani, G., Emerging links between type 2 diabetes and Alzheimer's disease. *World J Diabetes* **2015,** *6* (5), 744-51.

261. Verdile, G.; Fuller, S. J.; Martins, R. N., The role of type 2 diabetes in neurodegeneration. *Neurobiol Dis* **2015,** *84*, 22-38.

262. Zhao, W. Q.; Townsend, M., Insulin resistance and amyloidogenesis as common molecular foundation for type 2 diabetes and Alzheimer's disease. *Biochim Biophys Acta* **2009,** *1792* (5), 482-96.

263. Yang, Y.; Song, W., Molecular links between Alzheimer's disease and diabetes mellitus. *Neuroscience* **2013,** *250*, 140-50.

264. Jackson, K.; Barisone, G. A.; Diaz, E.; Jin, L. W.; DeCarli, C.; Despa, F., Amylin deposition in the brain: A second amyloid in Alzheimer disease? *Ann Neurol* **2013,** *74* (4), 517-26.

265. Hu, G.; Jousilahti, P.; Bidel, S.; Antikainen, R.; Tuomilehto, J., Type 2 diabetes and the risk of Parkinson's disease. *Diabetes Care* **2007,** *30* (4), 842-7.

266. Sandyk, R., The relationship between diabetes mellitus and Parkinson's disease. *Int J Neurosci* **1993,** *69* (1-4), 125-30.

267. Schernhammer, E.; Hansen, J.; Rugbjerg, K.; Wermuth, L.; Ritz, B., Diabetes and the risk of developing Parkinson's disease in Denmark. *Diabetes Care* **2011,** *34* (5), 1102-8.

268. Camargo Maluf, F.; Feder, D.; Alves de Siqueira Carvalho, A., Analysis of the Relationship between Type II Diabetes Mellitus and Parkinson's Disease: A Systematic Review. *Parkinsons Dis* **2019,** *2019*, 4951379.

269. Guo, J. L.; Covell, D. J.; Daniels, J. P.; Iba, M.; Stieber, A.; Zhang, B.; Riddle, D. M.; Kwong, L. K.; Xu, Y.; Trojanowski, J. Q.; Lee, V. M., Distinct alpha-synuclein strains differentially promote tau inclusions in neurons. *Cell* **2013,** *154* (1), 103-17.

270. Horvath, I.; Wittung-Stafshede, P., Cross-talk between amyloidogenic proteins in type-2 diabetes and Parkinson's disease. *Proc Natl Acad Sci U S A* **2016,** *113* (44), 12473-12477.

271. Mittal, K.; Mani, R. J.; Katare, D. P., Type 3 Diabetes: Cross Talk between Differentially Regulated Proteins of Type 2 Diabetes Mellitus and Alzheimer's Disease. *Sci Rep* **2016,** *6*, 25589.

272. Roberts, H. L.; Schneider, B. L.; Brown, D. R., alpha-Synuclein increases beta-amyloid secretion by promoting beta-/gamma-secretase processing of APP. *PLoS ONE* **2017,** *12* (2), e0171925.

273. Tsigelny, I. F.; Crews, L.; Desplats, P.; Shaked, G. M.; Sharikov, Y.; Mizuno, H.; Spencer, B.; Rockenstein, E.; Trejo, M.; Platoshyn, O.; Yuan, J. X.; Masliah, E., Mechanisms of hybrid oligomer formation in the pathogenesis of combined Alzheimer's and Parkinson's diseases. *PLoS ONE* **2008,** *3* (9), e3135.

274. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **2008,** *29* (11), 1859-65.

275. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L., 3rd; MacKerell, A. D., Jr.; Klauda, J. B.; Im, W., CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **2016,** *12* (1), 405-13.

276. Kieslich, C. A.; Tamamis, P.; Guzman, Y. A.; Onel, M.; Floudas, C. A., Highly Accurate Structure-Based Prediction of HIV-1 Coreceptor Usage Suggests Intermolecular Interactions Driving Tropism. *PLoS ONE* **2016,** *11* (2), e0148974.

277. Nilsson, L., Efficient table lookup without inverse square roots for calculation of pair wise atomic interactions in classical simulations. *J Comput Chem* **2009,** *30* (9), 1490-8.

278. Elber, R.; Ruymgaart, A. P.; Hess, B., SHAKE parallelization. *Eur Phys J Spec Top* **2011,** *200* (1), 211-223.

279. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977,** *23* (3), 327-341.

280. Lopez de Victoria, A.; Tamamis, P.; Kieslich, C. A.; Morikis, D., Insights into the structure, correlated motions, and electrostatic properties of two HIV-1 gp120 V3 loops. *PLoS ONE* **2012,** *7* (11), e49925.

281. Cheng, K. K.; Yeung, C. F.; Ho, S. W.; Chow, S. F.; Chow, A. H.; Baum, L., Highly stabilized curcumin nanoparticles tested in an in vitro blood-brain barrier model and in Alzheimer's disease Tg2576 mice. *AAPS J* **2013,** *15* (2), 324-36.

282. Kim, G. Y.; Kim, K. H.; Lee, S. H.; Yoon, M. S.; Lee, H. J.; Moon, D. O.; Lee, C. M.; Ahn, S. C.; Park, Y. C.; Park, Y. M., Curcumin inhibits immunostimulatory function of dendritic cells: MAPKs and translocation of NF-kappa B as potential targets. *J Immunol* **2005,** *174* (12), 8116-24.

283. Yang, F.; Lim, G. P.; Begum, A. N.; Ubeda, O. J.; Simmons, M. R.; Ambegaokar, S. S.; Chen, P. P.; Kayed, R.; Glabe, C. G.; Frautschy, S. A.; Cole, G. M., Curcumin inhibits formation of amyloid beta oligomers and fibrils, binds plaques, and reduces amyloid in vivo. *J Biol Chem* **2005,** *280* (7), 5892-901.

284. Mythri, R. B.; Bharath, M. M., Curcumin: a potential neuroprotective agent in Parkinson's disease. *Curr Pharm Des* **2012,** *18* (1), 91-9.

285. Nguyen, T. T.; Vuu, M. D.; Huynh, M. A.; Yamaguchi, M.; Tran, L. T.; Dang, T. P. T., Curcumin Effectively Rescued Parkinson's Disease-Like Phenotypes in a Novel Drosophila melanogaster Model with dUCH Knockdown. *Oxid Med Cell Longev* **2018,** *2018*, 2038267.

286. Sharma, N.; Nehru, B., Curcumin affords neuroprotection and inhibits alpha-synuclein aggregation in lipopolysaccharide-induced Parkinson's disease model. *Inflammopharmacology* **2018,** *26* (2), 349-360.

287. Pivari, F.; Mingione, A.; Brasacchio, C.; Soldati, L., Curcumin and Type 2 Diabetes Mellitus: Prevention and Treatment. *Nutrients* **2019,** *11* (8).

288. Thota, R. N.; Acharya, S. H.; Abbott, K. A.; Garg, M. L., Curcumin and long-chain Omega-3 polyunsaturated fatty acids for Prevention of type 2 Diabetes (COP-D): study protocol for a randomised controlled trial. *Trials* **2016,** *17* (1), 565.

289. Thota, R. N.; Rosato, J. I.; Dias, C. B.; Burrows, T. L.; Martins, R. N.; Garg, M. L., Dietary Supplementation with Curcumin Reduce Circulating Levels of Glycogen Synthase Kinase-3beta and Islet Amyloid Polypeptide in Adults with High Risk of Type 2 Diabetes and Alzheimer's Disease. *Nutrients* **2020,** *12* (4).

290. Thapa, A.; Jett, S. D.; Chi, E. Y., Curcumin Attenuates Amyloid-beta Aggregate Toxicity and Modulates Amyloid-beta Aggregation Pathway. *ACS Chem Neurosci* **2016,** *7* (1), 56-68.

291. Lo Cascio, F.; Puangmalai, N.; Ellsworth, A.; Bucchieri, F.; Pace, A.; Palumbo Piccionello, A.; Kayed, R., Toxic Tau Oligomers Modulated by Novel Curcumin Derivatives. *Sci Rep* **2019,** *9* (1), 19011.

292. Rane, J. S.; Bhaumik, P.; Panda, D., Curcumin Inhibits Tau Aggregation and Disintegrates Preformed Tau Filaments in vitro. *J Alzheimers Dis* **2017,** *60* (3), 999-1014.

293. Singh, P. K.; Kotia, V.; Ghosh, D.; Mohite, G. M.; Kumar, A.; Maji, S. K., Curcumin modulates alpha-synuclein aggregation and toxicity. *ACS Chem Neurosci* **2013,** *4* (3), 393-407.

294. Liu, G.; Gaines, J. C.; Robbins, K. J.; Lazo, N. D., Kinetic profile of amyloid formation in the presence of an aromatic inhibitor by nuclear magnetic resonance. *ACS Med Chem Lett* **2012,** *3* (10), 856-9.

295. Sparks, S.; Liu, G.; Robbins, K. J.; Lazo, N. D., Curcumin modulates the self-assembly of the islet amyloid polypeptide by disassembling alpha-helix. *Biochem Biophys Res Commun* **2012,** *422* (4), 551-5.

296. Jakubowski, J. M.; Orr, A. A.; Le, D. A.; Tamamis, P., Interactions between Curcumin Derivatives and Amyloid-beta Fibrils: Insights from Molecular Dynamics Simulations. *J Chem Inf Model* **2020,** *60* (1), 289-305.

297. Pierce, B. G.; Hourai, Y.; Weng, Z., Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS ONE* **2011,** *6* (9), e24657.

298. Masuda, Y.; Fukuchi, M.; Yatagawa, T.; Tada, M.; Takeda, K.; Irie, K.; Akagi, K.; Monobe, Y.; Imazawa, T.; Takegoshi, K., Solid-state NMR analysis of interaction sites of curcumin and 42-residue amyloid beta-protein fibrils. *Bioorg Med Chem* **2011,** *19* (20), 5967-74.

299. Carretero, M. I., Clay minerals and their beneficial effects upon human health. A review. *Appl Clay Sci* **2002,** *21* (3-4), 155-163.

300. Forrest, R. D., Early history of wound treatment. *J R Soc Med* **1982,** *75* (3), 198-205.

301. Pusch, R., *Bentonite Clay*. 1 ed.; Crc Press: Boca Raton, 2015.

302. Bergaya, F.; Theng, B. k. g.; Lagaly, G., *Handbook Of Clay Science, Volume 1 (developments In Clay Science)*. 1 ed.; Elsevier Science: 2006; p 1246.

303. Moraes, J. D. D.; Bertolino, S. R. A.; Cuffini, S. L.; Ducart, D. F.; Bretzke, P. E.; Leonardi, G. R., Clay minerals: Properties and applications to dermocosmetic products and perspectives of natural raw materials for therapeutic purposes-A review. *Int J Pharm* **2017,** *534* (1-2), 213-219.

304. Rafati, Z.; Sirousazar, M.; Hassan, Z. M.; Kheiri, F., Honey-Loaded Egg White/Poly(vinyl alcohol)/Clay Bionanocomposite Hydrogel Wound Dressings: In Vitro and In Vivo Evaluations. *J Polym Environ* **2019,** *28* (1), 32-46.

305. Saary, J.; Qureshi, R.; Palda, V.; DeKoven, J.; Pratt, M.; Skotnicki-Grant, S.; Holness, L., A systematic review of contact dermatitis treatment and prevention. *J Am Acad Dermatol* **2005,** *53* (5), 845.

306. Saha, K.; Dutta, K.; Basu, A.; Adhikari, A.; Chattopadhyay, D.; Sarkar, P., Controlled delivery of tetracycline hydrochloride intercalated into smectite clay using polyurethane nanofibrous membrane for wound healing application. *Nano-Structures & Nano-Objects* **2020,** *21*, 100418.

307. Viseras, C.; Carazo, E.; Borrego-Sanchez, A.; Garcia-Villen, F.; Sanchez-Espejo, R.; Cerezo, P.; Aguzzi, C., Clay Minerals in Skin Drug Delivery. *Clays Clay Miner* **2019,** *67* (1), 59-71.

308. Lee, Y. H.; Chen, B. Y.; Lin, F. H.; Lin, K. Y.; Lin, K. F., Cytotoxic Assessment of L-Ascorbic Acid/Montmorillonite Upon Human Dermal Fibroblasts in Vitro: Mtt Activity Assay. *Biomedical Engineering-Applications Basis Communications* **2008,** *20* (6), 337-343.

309. McMahan, Z. H.; DuPont, H. L., Review article: the history of acute infectious diarrhoea management--from poorly focused empiricism to fluid therapy and modern pharmacotherapy. *Aliment Pharmacol Ther* **2007,** *25* (7), 759-69.

310. Ping, Y.; Hu, X.; Yao, Q.; Hu, Q.; Amini, S.; Miserez, A.; Tang, G., Engineering bioinspired bacteria-adhesive clay nanoparticles with a membrane-

disruptive property for the treatment of Helicobacter pylori infection. *Nanoscale* **2016,** *8* (36), 16486-98.

311. Walker, R. R., Kaolin in the Treatment of Asiatic Cholera: its Action and Uses. *Proc R Soc Med* **1921,** *14* (Ther Pharmacol Sect), 23-9.

312. Williams, L. B.; Haydel, S. E., Evaluation of the medicinal use of clay minerals as antibacterial agents. *Int Geol Rev* **2010,** *52* (7/8), 745-770.

313. Chang, F. Y.; Lu, C. L.; Chen, C. Y.; Luo, J. C., Efficacy of dioctahedral smectite in treating patients of diarrhea-predominant irritable bowel syndrome. *J Gastroenterol Hepatol* **2007,** *22* (12), 2266-72.

314. Phillips, T. D.; Wang, M.; Elmore, S. E.; Hearon, S.; Wang, J. S., NovaSil clay for the protection of humans and animals from aflatoxins and other contaminants. *Clays Clay Miner* **2019,** *67* (1), 99-110.

315. Nwosu, F. O.; Ajala, O. J.; Owoyemi, R. M.; Raheem, B. G., Preparation and characterization of adsorbents derived from bentonite and kaolin clays. *Appl Water Sci* **2018,** *8* (7), 195.

316. Williams, L. B.; Metge, D. W.; Eberl, D. D.; Harvey, R. W.; Turner, A. G.; Prapaipong, P.; Poret-Peterson, A. T., What makes a natural clay antibacterial? *Environ Sci Technol* **2011,** *45* (8), 3768-73.

317. Alamudy, H. A.; Cho, K., Selective adsorption of cesium from an aqueous solution by a montmorillonite-prussian blue hybrid. *Chemical Engineering Journal* **2018,** *349*, 595-602.

318. Ntwampe, O. I.; Moothi, K., Removal of Heavy Metals Using Bentonite Clay and Inorganic Coagulants. In *Heavy Metals*, Saleh, H. E.-D. M.; Aglan, R. F., Eds. InTech: 2018.

319. Wang, M.; Phillips, T. D., Potential Applications of Clay-Based Therapy for the Reduction of Pesticide Exposures in Humans and Animals. *Appl Sci (Basel)* **2019,** *9* (24), 5325.

320. Wang, M.; Safe, S.; Hearon, S. E.; Phillips, T. D., Strong adsorption of Polychlorinated Biphenyls by processed montmorillonite clays: Potential applications as toxin enterosorbents during disasters and floods. *Environ Pollut* **2019,** *255* (Pt 1), 113210.

321. Grant, P. G.; Phillips, T. D., Isothermal Adsorption of Aflatoxin B(1) on HSCAS Clay. *J Agric Food Chem* **1998,** *46* (2), 599-605.

322. Brown, K. A.; Mays, T.; Romoser, A.; Marroquin-Cardona, A.; Mitchell, N. J.; Elmore, S. E.; Phillips, T. D., Modified hydra bioassay to evaluate the toxicity of multiple mycotoxins and predict the detoxification efficacy of a clay-based sorbent. *J Appl Toxicol* **2014,** *34* (1), 40-8.

323. Hearon, S. E.; Wang, M.; Phillips, T. D., Strong Adsorption of Dieldrin by Parent and Processed Montmorillonite Clays. *Environ Toxicol Chem* **2020,** *39* (3), 517-525.

324. Wang, M.; Maki, C. R.; Deng, Y.; Tian, Y.; Phillips, T. D., Development of High Capacity Enterosorbents for Aflatoxin B1 and Other Hazardous Chemicals. *Chem Res Toxicol* **2017,** *30* (9), 1694-1701.

325. Mitchell, N. J.; Kumi, J.; Aleser, M.; Elmore, S. E.; Rychlik, K. A.; Zychowski, K. E.; Romoser, A. A.; Phillips, T. D.; Ankrah, N. A., Short-term safety and efficacy of calcium montmorillonite clay (UPSN) in children. *Am J Trop Med Hyg* **2014,** *91* (4), 777-85.

326. Awuor, A. O.; Yard, E.; Daniel, J. H.; Martin, C.; Bii, C.; Romoser, A.; Oyugi, E.; Elmore, S.; Amwayi, S.; Vulule, J.; Zitomer, N. C.; Rybak, M. E.; Phillips, T. D.; Montgomery, J. M.; Lewis, L. S., Evaluation of the efficacy, acceptability and palatability of calcium montmorillonite clay used to reduce aflatoxin B1 dietary exposure in a crossover study in Kenya. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* **2017,** *34* (1), 93-102.

327. Pollock, B. H.; Elmore, S.; Romoser, A.; Tang, L.; Kang, M. S.; Xue, K.; Rodriguez, M.; Dierschke, N. A.; Hayes, H. G.; Hansen, H. A.; Guerra, F.; Wang, J. S.; Phillips, T., Intervention trial with calcium montmorillonite clay in a south Texas population exposed to aflatoxin. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* **2016,** *33* (8), 1346-54.

328. Maki, C. R.; Monteiro, A. P. A.; Elmore, S. E.; Tao, S.; Bernard, J. K.; Harvey, R. B.; Romoser, A. A.; Phillips, T. D., Calcium montmorillonite clay in dairy feed reduces aflatoxin concentrations in milk without interfering with milk quality, composition or yield. *Anim. Feed Sci. Technol.* **2016,** *214*, 130-135.

329. Maki, C. R.; Thomas, A. D.; Elmore, S. E.; Romoser, A. A.; Harvey, R. B.; Ramirez-Ramirez, H. A.; Phillips, T. D., Effects of calcium montmorillonite clay and aflatoxin exposure on dry matter intake, milk production, and milk composition. *J Dairy Sci* **2016,** *99* (2), 1039-1046.

330. Mitchell, N. J.; Xue, K. S.; Lin, S.; Marroquin-Cardona, A.; Brown, K. A.; Elmore, S. E.; Tang, L.; Romoser, A.; Gelderblom, W. C.; Wang, J. S.; Phillips, T. D., Calcium montmorillonite clay reduces AFB1 and FB1 biomarkers in rats exposed to

single and co-exposures of aflatoxin and fumonisin. *J Appl Toxicol* **2014,** *34* (7), 795-804.

331. Phillips, T. D., Calcium Montmorillonite Clay for the Reduction of Aflatoxin Residues in Milk and Dairy Products. *JDVS* **2017,** *2* (3).

332. Fukushima, Y., X-Ray-Diffraction Study of Aqueous Montmorillonite Emulsions. *Clays Clay Miner* **1984,** *32* (4), 320-326.

333. Madejova, J.; Sekerakova, L.; Bizovska, V.; Slany, M.; Jankovic, L., Near-infrared spectroscopy as an effective tool for monitoring the conformation of alkylammonium surfactants in montmorillonite interlayers. *Vib Spectrosc* **2016,** *84*, 44-52.

334. Fajnor, V. S.; Jesenak, K., Differential thermal analysis of montmorillonite. *Journal of Thermal Analysis and Calorimetry* **1996,** *46* (2), 489-493.

335. P, V.; L, T., Thermo-Analytical Techniques on MX-80 Montmorillonite: A Way to Know the Behavior of Water and its Thermodynamic Properties during Hydration – Dehydration Processes. *Pharm Anal Acta* **2015,** *07* (02).

336. Fitaroni, L. B.; de Lima, J. A.; Cruz, S. A.; Waldman, W. R., Thermal stability of polypropylene-montmorillonite clay nanocomposites: Limitation of the thermogravimetric analysis. *Polym Degrad Stab* **2015,** *111*, 102-108.

337. Drummy, L. F.; Koerner, H.; Farmer, K.; Tan, A.; Farmer, B. L.; Vaia, R. A., High-resolution electron microscopy of montmorillonite and montmorillonite/epoxy nanocomposites. *J Phys Chem B* **2005,** *109* (38), 17868-78.

338. Takahashi, C.; Shirai, T.; Fuji, M., Electron microscopic observation of montmorillonite swelled by water with the aid of hydrophilic ionic liquid. *Materials Chemistry and Physics* **2013,** *141* (2-3), 657-664.

339. del Mar Orta, M.; Martín, J.; Medina-Carrasco, S.; Santos, J. L.; Aparicio, I.; Alonso, E., Adsorption of propranolol onto montmorillonite: Kinetic, isotherm and pH studies. *Appl Clay Sci* **2019,** *173*, 107-114.

340. Ayawei, N.; Ebelegi, A. N.; Wankasi, D., Modelling and Interpretation of Adsorption Isotherms. *J Chem* **2017,** *2017*, 1-11.

341. Awad, A. M.; Shaikh, S. M. R.; Jalab, R.; Gulied, M. H.; Nasser, M. S.; Benamor, A.; Adham, S., Adsorption of organic pollutants by natural and modified clays: A comprehensive review. *Separation and Purification Technology* **2019,** *228*, 115719.

342. Elliott, C. T.; Connolly, L.; Kolawole, O., Potential adverse effects on animal health and performance caused by the addition of mineral adsorbents to feeds to reduce mycotoxin exposure. *Mycotoxin Res* **2020,** *36* (1), 115-126.

343. Bickmore, B. R.; Rosso, K. M.; Nagy, K. L.; Cygan, R. T.; Tadanier, C. J., Ab initio determination of edge surface structures for dioctahedral 2 : 1 phyllosilicates: Implications for acid-base reactivity. *Clays Clay Miner* **2003,** *51* (4), 359-371.

344. Geatches, D. L.; Jacquet, A.; Clark, S. J.; Greenwell, H. C., Monomer Adsorption on Kaolinite: Modeling the Essential Ingredients. *Journal of Physical Chemistry C* **2012,** *116* (42), 22365-22374.

345. Sánchez, V. M.; Miranda, C. R., Modeling Acid Oil Component Interactions with Carbonate Reservoirs: A First-Principles View on Low Salinity Recovery Mechanisms. *J. Phys. Chem. C* **2014,** *118* (33), 19180-19187.

346. Zhong, J.; Wang, X.; Du, J.; Wang, L.; Yan, Y.; Zhang, J., Combined Molecular Dynamics and Quantum Mechanics Study of Oil Droplet Adsorption on Different Self-Assembly Monolayers in Aqueous Solution. *J. Phys. Chem. C* **2013,** *117* (24), 12510-12519.

347. Underwood, T.; Erastova, V.; Cubillas, P.; Greenwell, H. C., Molecular Dynamic Simulations of Montmorillonite Organic Interactions under Varying Salinity: An Insight into Enhanced Oil Recovery. *Journal of Physical Chemistry C* **2015,** *119* (13), 7282-7294.

348. Aggarwal, V.; Chien, Y. Y.; Teppen, B. J., Molecular simulations to estimate thermodynamics for adsorption of polar organic solutes to montmorillonite. *Eur J Soil Sci* **2007,** *58* (4), 945-957.

349. Heinz, H.; Vaia, R. A.; Krishnamoorti, R.; Farmer, B. L., Self-assembly of alkylammonium chains on montmorillonite: Effect of chain length, head group structure, and cation exchange capacity. *Chem. Mater.* **2007,** *19* (1), 59-68.

350. Kristóf, T.; Sarkadi, Z.; Ható, Z.; Rutkai, G., Simulation study of intercalation complexes of kaolinite with simple amides as primary intercalation reagents. *Comp Mater Sci* **2018,** *143*, 118-125.

351. Liu, C.; Gu, C.; Yu, K.; Li, H.; Teppen, B. J.; Johnston, C. T.; Boyd, S. A.; Zhou, D., Integrating structural and thermodynamic mechanisms for sorption of PCBs by montmorillonite. *Environ Sci Technol* **2015,** *49* (5), 2796-805.

352. Samaraweera, M.; Jolin, W.; Vasudevan, D.; MacKay, A. A.; Gascon, J. A., Atomistic Prediction of Sorption Free Energies of Cationic Aromatic Amines on

Montmorillonite: A Linear Interaction Energy Method. *Environ Sci Technol Lett* **2014,** *1* (6), 284-289.

353. Willemsen, J. A. R.; Bourg, I. C., Molecular dynamics simulation of the adsorption of per- and polyfluoroalkyl substances (PFASs) on smectite clay. *J Colloid Interface Sci* **2021,** *585*, 337-346.

354. Heinz, H.; Lin, T. J.; Mishra, R. K.; Emami, F. S., Thermodynamically consistent force fields for the assembly of inorganic, organic, and biological nanostructures: the INTERFACE force field. *Langmuir* **2013,** *29* (6), 1754-65.

355. Heinz, H.; Koerner, H.; Anderson, K. L.; Vaia, R. A.; Farmer, B. L., Force field for mica-type silicates and dynamics of octadecylammonium chains grafted to montmorillonite. *Chem. Mater.* **2005,** *17* (23), 5658-5669.

356. Abramyan, T. M.; Hyde-Volpe, D. L.; Stuart, S. J.; Latour, R. A., Application of advanced sampling and analysis methods to predict the structure of adsorbed protein on a material surface. *Biointerphases* **2017,** *12* (2), 02D409.

357. Emami, F. S.; Puddu, V.; Berry, R. J.; Varshney, V.; Patwardhan, S. V.; Perry, C. C.; Heinz, H., Prediction of Specific Biomolecule Adsorption on Silica Surfaces as a Function of pH and Particle Size. *Chem. Mater.* **2014,** *26* (19), 5725-5734.

358. Heinz, H.; Vaia, R. A.; Farmer, B. L., Interaction energy and surface reconstruction between sheets of layered silicates. *J Chem Phys* **2006,** *124* (22), 224713.

359. Latour, R. A., 3.14 Molecular Simulation Methods to Investigate Protein Adsorption Behavior at the Atomic Level ☆. In *Comprehensive Biomaterials II*, Elsevier: 2017; pp 268-294.

360. Mako, E.; Kovacs, A.; Hato, Z.; Zsirka, B.; Kristof, T., Characterization of kaolinite-ammonium acetate complexes prepared by one-step homogenization method. *J Colloid Interface Sci* **2014,** *431*, 125-31.

361. Ramakrishnan, S. K.; Zhu, J.; Gergely, C., Organic-inorganic interface simulation for new material discoveries. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2017,** *7* (1), e1277.

362. Greenland, D. J., Adsorption of 1-n-Alkyl Pyridinium Bromides by Montmorillonite. *Clays Clay Miner* **1960,** *9* (1), 484-499.

363. Sterling, T.; Irwin, J. J., ZINC 15--Ligand Discovery for Everyone. *J Chem Inf Model* **2015,** *55* (11), 2324-37.

364. Bailey, G. W.; White, J. L.; Rothberg, T., Adsorption of Organic Herbicides by Montmorillonite - Role of Ph and Chemical Character of Adsorbate. *Soil Science Society of America Proceedings* **1968,** *32* (2), 222-&.

365. Herwig, U.; Klumpp, E.; Narres, H. D.; Schwuger, M. J., Physicochemical interactions between atrazine and clay minerals. *Appl Clay Sci* **2001,** *18* (5-6), 211-222.

366. Burns, D. C.; Ellis, D. A.; Li, H.; McMurdo, C. J.; Webster, E., Experimental pKa Determination for Perfluorooctanoic Acid (PFOA) and the Potential Impact of pKa Concentration Dependence on Laboratory-Measured Partitioning Phenomena and Environmental Modeling. *Environ Sci Technol* **2008,** *42* (24), 9283-9288.

367. Steinle-Darling, E.; Reinhard, M., Nanofiltration for Trace Organic Contaminant Removal: Structure, Solution, and Membrane Fouling Effects on the Rejection of Perfluorochemicals. *Environ Sci Technol* **2008,** *42* (14), 5292-5297.

368. Vierke, L.; Berger, U.; Cousins, I. T., Estimation of the Acid Dissociation Constant of Perfluoroalkyl Carboxylic Acids through an Experimental Investigation of their Water-to-Air Transport. *Environ Sci Technol* **2013,** *47* (19), 11032-11039.

369. Fortenberry, G. Z.; Beckman, J.; Schwartz, A.; Prado, J. B.; Graham, L. S.; Higgins, S.; Lackovic, M.; Mulay, P.; Bojes, H.; Waltz, J.; Mitchell, Y.; Leinenkugel, K.; Oriel, M. S.; Evans, E.; Calvert, G. M., Magnitude and characteristics of acute paraquat- and diquat-related illnesses in the US: 1998-2013. *Environ Res* **2016,** *146*, 191-9.

370. Mertens, M.; Hoss, S.; Neumann, G.; Afzal, J.; Reichenbecher, W., Glyphosate, a chelating agent-relevant for ecological risk assessment? *Environ Sci Pollut Res Int* **2018,** *25* (6), 5298-5317.

371. Mesnage, R.; Antoniou, M. N., Facts and Fallacies in the Debate on Glyphosate Toxicity. *Front Public Health* **2017,** *5*, 316.

372. Samsel, A.; Seneff, S., Glyphosate's Suppression of Cytochrome P450 Enzymes and Amino Acid Biosynthesis by the Gut Microbiome: Pathways to Modern Diseases. *Entropy* **2013,** *15* (4), 1416-1463.

373. Samsel, A.; Seneff, S., Glyphosate, pathways to modern diseases II: Celiac sprue and gluten intolerance. *Interdiscip Toxicol* **2013,** *6* (4), 159-84.

374. Samsel, A.; Seneff, S., Glyphosate, pathways to modern diseases III: Manganese, neurological diseases, and associated pathologies. *Surg Neurol Int* **2015,** *6*, 45.

375. Seneff, S.; Samsel, A., Glyphosate, pathways to modern diseases IV: cancer and related pathologies. *Journal of Biological Physics and Chemistry* **2015,** *15* (3), 121-159.

376. Seneff, S.; Samsel, A., Glyphosate pathways to modern diseases V: Amino acid analogue of glycine in diverse proteins. *Journal of Biological Physics and Chemistry* **2016,** *16* (1), 9-46.

377. Myers, J. P.; Antoniou, M. N.; Blumberg, B.; Carroll, L.; Colborn, T.; Everett, L. G.; Hansen, M.; Landrigan, P. J.; Lanphear, B. P.; Mesnage, R.; Vandenberg, L. N.; Vom Saal, F. S.; Welshons, W. V.; Benbrook, C. M., Concerns over use of glyphosate-based herbicides and risks associated with exposures: a consensus statement. *Environ Health* **2016,** *15*, 19.

378. Vandenberg, L. N.; Blumberg, B.; Antoniou, M. N.; Benbrook, C. M.; Carroll, L.; Colborn, T.; Everett, L. G.; Hansen, M.; Landrigan, P. J.; Lanphear, B. P.; Mesnage, R.; Vom Saal, F. S.; Welshons, W. V.; Myers, J. P., Is it time to reassess current safety standards for glyphosate-based herbicides? *J Epidemiol Community Health* **2017,** *71* (6), 613-618.

379. Mesnage, R.; Defarge, N.; Spiroux de Vendomois, J.; Seralini, G. E., Potential toxic effects of glyphosate and its commercial formulations below regulatory limits. *Food Chem Toxicol* **2015,** *84*, 133-53.

380. Wesseling, C.; van Wendel de Joode, B.; Ruepert, C.; Leon, C.; Monge, P.; Hermosillo, H.; Partanen, T. J., Paraquat in developing countries. *Int J Occup Environ Health* **2001,** *7* (4), 275-86.

381. Park, S. K.; Kang, D.; Beane-Freeman, L.; Blair, A.; Hoppin, J. A.; Sandler, D. P.; Lynch, C. F.; Knott, C.; Gwak, J.; Alavanja, M., Cancer incidence among paraquat exposed applicators in the agricultural health study: prospective cohort study. *Int J Occup Environ Health* **2009,** *15* (3), 274-81.

382. Dinis-Oliveira, R. J.; Duarte, J. A.; Sanchez-Navarro, A.; Remiao, F.; Bastos, M. L.; Carvalho, F., Paraquat poisonings: mechanisms of lung toxicity, clinical features, and treatment. *Crit Rev Toxicol* **2008,** *38* (1), 13-71.

383. Engel, L. S.; Hill, D. A.; Hoppin, J. A.; Lubin, J. H.; Lynch, C. F.; Pierce, J.; Samanic, C.; Sandler, D. P.; Blair, A.; Alavanja, M. C., Pesticide use and breast cancer risk among farmers' wives in the agricultural health study. *Am J Epidemiol* **2005,** *161* (2), 121-35.

384. Kamel, F.; Tanner, C.; Umbach, D.; Hoppin, J.; Alavanja, M.; Blair, A.; Comyns, K.; Goldman, S.; Korell, M.; Langston, J.; Ross, G.; Sandler, D., Pesticide exposure and self-reported Parkinson's disease in the agricultural health study. *Am J Epidemiol* **2007,** *165* (4), 364-74.

385. Tanner, C. M.; Kamel, F.; Ross, G. W.; Hoppin, J. A.; Goldman, S. M.; Korell, M.; Marras, C.; Bhudhikanok, G. S.; Kasten, M.; Chade, A. R.; Comyns, K.;

Richards, M. B.; Meng, C.; Priestley, B.; Fernandez, H. H.; Cambi, F.; Umbach, D. M.; Blair, A.; Sandler, D. P.; Langston, J. W., Rotenone, paraquat, and Parkinson's disease. *Environ Health Perspect* **2011,** *119* (6), 866-72.

386. Euripidou, E.; Murray, V., Public health impacts of floods and chemical contamination. *Journal of Public Health* **2004,** *26* (4), 376-383.

387. Glass, R. L., Adsorption of Glyphosate by Soils and Clay-Minerals. *J Agric Food Chem* **1987,** *35* (4), 497-500.

388. Munira, S.; Farenhorst, A.; Akinremi, W., Phosphate and glyphosate sorption in soils following long-term phosphate applications. *Geoderma* **2018,** *313*, 146-153.

389. Zen, J. M.; Jeng, S. H.; Chen, H. J., Determination of paraquat by square-wave voltammetry at a perfluorosulfonated ionomer/clay-modified electrode. *Anal Chem* **1996,** *68* (3), 498-502.

390. Chou, K.; Wright, R. O., Phthalates in food and medical devices. *J Med Toxicol* **2006,** *2* (3), 126-35.

391. Lioy, P. J.; Hauser, R.; Gennings, C.; Koch, H. M.; Mirkes, P. E.; Schwetz, B. A.; Kortenkamp, A., Assessment of phthalates/phthalate alternatives in children's toys and childcare articles: Review of the report including conclusions and recommendation of the Chronic Hazard Advisory Panel of the Consumer Product Safety Commission. *J Expo Sci Environ Epidemiol* **2015,** *25* (4), 343-53.

392. North, E. J.; Halden, R. U., Plastics and environmental health: the road ahead. *Rev Environ Health* **2013,** *28* (1), 1-8.

393. Kubwabo, C.; Kosarac, I.; Stewart, B.; Gauthier, B. R.; Lalonde, K.; Lalonde, P. J., Migration of bisphenol A from plastic baby bottles, baby bottle liners and reusable polycarbonate drinking bottles. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* **2009,** *26* (6), 928-37.

394. Rosenmai, A. K.; Dybdahl, M.; Pedersen, M.; Alice van Vugt-Lussenburg, B. M.; Wedebye, E. B.; Taxvig, C.; Vinggaard, A. M., Are structural analogues to bisphenol a safe alternatives? *Toxicol Sci* **2014,** *139* (1), 35-47.

395. Feng, Y.; Jiao, Z.; Shi, J.; Li, M.; Guo, Q.; Shao, B., Effects of bisphenol analogues on steroidogenic gene expression and hormone synthesis in H295R cells. *Chemosphere* **2016,** *147*, 9-19.

396. Konieczna, A.; Rutkowska, A.; Rachon, D., Health risk of exposure to Bisphenol A (BPA). *Rocz Panstw Zakl Hig* **2015,** *66* (1), 5-11.

397. Moon, M. K., Concern about the Safety of Bisphenol A Substitutes. *Diabetes Metab J* **2019,** *43* (1), 46-48.

398. Pjanic, M., The role of polycarbonate monomer bisphenol-A in insulin resistance. *PeerJ* **2017,** *5*, e3809.

399. Talsness, C. E.; Andrade, A. J.; Kuriyama, S. N.; Taylor, J. A.; vom Saal, F. S., Components of plastic: experimental studies in animals and relevance for human health. *Philos Trans R Soc Lond B Biol Sci* **2009,** *364* (1526), 2079-96.

400. Tarapore, P.; Ying, J.; Ouyang, B.; Burke, B.; Bracken, B.; Ho, S. M., Exposure to bisphenol A correlates with early-onset prostate cancer and promotes centrosome amplification and anchorage-independent growth in vitro. *PLoS ONE* **2014,** *9* (3), e90332.

401. Ullah, H.; Jahan, S.; Ain, Q. U.; Shaheen, G.; Ahsan, N., Effect of bisphenol S exposure on male reproductive system of rats: A histological and biochemical study. *Chemosphere* **2016,** *152*, 383-91.

402. Akingbemi, B. T.; Ge, R.; Klinefelter, G. R.; Zirkin, B. R.; Hardy, M. P., Phthalate-induced Leydig cell hyperplasia is associated with multiple endocrine disturbances. *Proc Natl Acad Sci U S A* **2004,** *101* (3), 775-80.

403. Gray, L. E., Jr.; Barlow, N. J.; Howdeshell, K. L.; Ostby, J. S.; Furr, J. R.; Gray, C. L., Transgenerational effects of Di (2-ethylhexyl) phthalate in the male CRL:CD(SD) rat: added value of assessing multiple offspring per litter. *Toxicol Sci* **2009,** *110* (2), 411-25.

404. Rusyn, I.; Peters, J. M.; Cunningham, M. L., Modes of action and species-specific effects of di-(2-ethylhexyl)phthalate in the liver. *Crit Rev Toxicol* **2006,** *36* (5), 459-79.

405. Xu, H.; Shao, X.; Zhang, Z.; Zou, Y.; Wu, X.; Yang, L., Oxidative stress and immune related gene expression following exposure to di-n-butyl phthalate and diethyl phthalate in zebrafish embryos. *Ecotoxicol Environ Saf* **2013,** *93*, 39-44.

406. Braun, J. M.; Sathyanarayana, S.; Hauser, R., Phthalate exposure and children's health. *Curr Opin Pediatr* **2013,** *25* (2), 247-54.

407. Grindler, N. M.; Vanderlinden, L.; Karthikraj, R.; Kannan, K.; Teal, S.; Polotsky, A. J.; Powell, T. L.; Yang, I. V.; Jansson, T., Exposure to Phthalate, an Endocrine Disrupting Chemical, Alters the First Trimester Placental Methylome and Transcriptome in Women. *Sci Rep* **2018,** *8* (1), 6086.

408. Jensen, M. S.; Norgaard-Pedersen, B.; Toft, G.; Hougaard, D. M.; Bonde, J. P.; Cohen, A.; Thulstrup, A. M.; Ivell, R.; Anand-Ivell, R.; Lindh, C. H.; Jonsson, B. A., Phthalates and perfluorooctanesulfonic acid in human amniotic fluid: temporal trends and timing of amniocentesis in pregnancy. *Environ Health Perspect* **2012,** *120* (6), 897-903.

409. Swan, S. H.; Main, K. M.; Liu, F.; Stewart, S. L.; Kruse, R. L.; Calafat, A. M.; Mao, C. S.; Redmon, J. B.; Ternand, C. L.; Sullivan, S.; Teague, J. L.; Study for Future Families Research, T., Decrease in anogenital distance among male infants with prenatal phthalate exposure. *Environ Health Perspect* **2005,** *113* (8), 1056-61.

410. Betts, K. S., Plastics and food sources: dietary intervention to reduce BPA and DEHP. *Environ Health Perspect* **2011,** *119* (7), A306.

411. Per and Polyfluoroalkyl Substances (PFAS) | FDA.

412. Gao, K.; Zhuang, T.; Liu, X.; Fu, J.; Zhang, J.; Fu, J.; Wang, L.; Zhang, A.; Liang, Y.; Song, M.; Jiang, G., Prenatal Exposure to Per- and Polyfluoroalkyl Substances (PFASs) and Association between the Placental Transfer Efficiencies and Dissociation Constant of Serum Proteins-PFAS Complexes. *Environ Sci Technol* **2019,** *53* (11), 6529-6538.

413. Wang, Z.; DeWitt, J. C.; Higgins, C. P.; Cousins, I. T., A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)? *Environ Sci Technol* **2017,** *51* (5), 2508-2518.

414. Aimuzi, R.; Luo, K.; Huang, R.; Huo, X.; Nian, M.; Ouyang, F.; Du, Y.; Feng, L.; Wang, W.; Zhang, J.; Shanghai Birth Cohort, S., Perfluoroalkyl and polyfluroalkyl substances and maternal thyroid hormones in early pregnancy. *Environ Pollut* **2020,** *264*, 114557.

415. Shih, K. M.; Wang, F., Adsorption behavior of perfluorochemicals (PFCs) on boehmite: influence of solution chemistry. *2013 International Symposium on Environmental Science and Technology (2013 Isest)* **2013,** *18*, 106-113.

416. Onel, M.; Beykal, B.; Wang, M.; Grimm, F. A.; Zhou, L.; Wright, F. A.; Phillips, T. D.; Rusyn, I.; Pistikopoulos, E. N., Optimal Chemical Grouping and Sorbent Material Design by Data Analysis, Modeling and Dimensionality Reduction Techniques. *ESCAPE* **2018,** *43*, 421-426.

417. Hearon, S. E.; Wang, M.; McDonald, T. J.; Phillips, T. D., Decreased bioavailability of aminomethylphosphonic acid (AMPA) in genetically modified corn with activated carbon or calcium montmorillonite clay inclusion in soil. *J Environ Sci (China)* **2021,** *100*, 131-143.

418. Wang, M.; Hearon, S. E.; Phillips, T. D., Development of enterosorbents that can be added to food and water to reduce toxin exposures during disasters. *J Environ Sci Health B* **2019,** *54* (6), 514-524.

419. Mallajosyula, S. S.; Guvench, O.; Hatcher, E.; Mackerell, A. D., Jr., CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. *J Chem Theory Comput* **2012,** *8* (2), 759-776.

420. Wang, Y.-J.; Li, C.-B.; Zhou, D.-M.; Friedman, S. P., Wien effect in suspensions and its application in soil science. Elsevier: 2013; Vol. 122, pp 127-178.

421. Wang, M.; Hearon, S. E.; Johnson, N. M.; Phillips, T. D., Development of broad-acting clays for the tight adsorption of benzo[a]pyrene and aldicarb. *Appl Clay Sci* **2019,** *168*, 196-202.

422. Shao, Q.; Hall, C. K., Protein adsorption on nanoparticles: model development using computer simulation. *J Phys Condens Matter* **2016,** *28* (41), 414019.

423. Shao, Q.; Hall, C. K., Binding Preferences of Amino Acids for Gold Nanoparticles: A Molecular Simulation Study. *Langmuir* **2016,** *32* (31), 7888-96.

424. Shao, Q.; Hall, C. K., Allosteric effects of gold nanoparticles on human serum albumin. *Nanoscale* **2017,** *9* (1), 380-390.

425. Zhao, X.; Pan, F.; Xu, H.; Yaseen, M.; Shan, H.; Hauser, C. A.; Zhang, S.; Lu, J. R., Molecular self-assembly and applications of designer peptide amphiphiles. *Chem Soc Rev* **2010,** *39* (9), 3480-98.

426. Shi, J.; Xu, B., Nanoscale Assemblies of Small Molecules Control the Fate of Cells. *Nano Today* **2015,** *10* (5), 615-630.

427. Tu, R. S.; Tirrell, M., Bottom-up design of biomimetic assemblies. *Adv Drug Deliv Rev* **2004,** *56* (11), 1537-63.

428. Sun, L.; Zheng, C.; Webster, T. J., Self-assembled peptide nanomaterials for biomedical applications: promises and pitfalls. *Int J Nanomedicine* **2017,** *12*, 73-86.

429. Panda, J. J.; Dua, R.; Mishra, A.; Mittra, B.; Chauhan, V. S., 3D cell growth and proliferation on a RGD functionalized nanofibrillar hydrogel based on a conformationally restricted residue containing dipeptide. *ACS Appl Mater Interfaces* **2010,** *2* (10), 2839-48.

430. Deidda, G.; Jonnalagadda, S. V. R.; Spies, J. W.; Ranella, A.; Mossou, E.; Forsyth, V. T.; Mitchell, E. P.; Bowler, M. W.; Tamamis, P.; Mitraki, A., Self-

Assembled Amyloid Peptides with Arg-Gly-Asp (RGD) Motifs As Scaffolds for Tissue Engineering. *ACS Biomater Sci Eng* **2017,** *3* (7), 1404-1416.

431. Jonnalagadda, S. V. R.; Ornithopoulou, E.; Orr, A. A.; Mossou, E.; Forsyth, V. T.; Mitchell, E. P.; Bowler, M. W.; Mitraki, A.; Tamamis, P., Computational design of amyloid self-assembling peptides bearing aromatic residues and the cell adhesive motif Arg-Gly-Asp. *Molecular Systems Design & Engineering* **2017,** *2* (3), 321-335.

432. Kokotidou, C.; Jonnalagadda, S. V. R.; Orr, A. A.; Seoane-Blanco, M.; Apostolidou, C. P.; van Raaij, M. J.; Kotzabasaki, M.; Chatzoudis, A.; Jakubowski, J. M.; Mossou, E.; Forsyth, V. T.; Mitchell, E. P.; Bowler, M. W.; Llamas-Saiz, A. L.; Tamamis, P.; Mitraki, A., A novel amyloid designable scaffold and potential inhibitor inspired by GAIIG of amyloid beta and the HIV-1 V3 loop. *FEBS Lett* **2018,** *592* (11), 1777-1788.

433. Kokotidou, C.; Jonnalagadda, S. V. R.; Orr, A. A.; Vrentzos, G.; Kretsovali, A.; Tamamis, P.; Mitraki, A. A., Designer Amyloid Cell-Penetrating Peptides for Potential Use as Gene Transfer Vehicles. *Biomolecules* **2019,** *10* (1).

434. Ghosh, M.; Halperin-Sternfeld, M.; Grigoriants, I.; Lee, J.; Nam, K. T.; Adler-Abramovich, L., Arginine-Presenting Peptide Hydrogels Decorated with Hydroxyapatite as Biomimetic Scaffolds for Bone Regeneration. *Biomacromolecules* **2017,** *18* (11), 3541-3550.

435. Ghosh, M.; Halperin-Sternfeld, M.; Grinberg, I.; Adler-Abramovich, L., Injectable Alginate-Peptide Composite Hydrogel as a Scaffold for Bone Tissue Regeneration. *Nanomaterials (Basel)* **2019,** *9* (4).

436. Loo, Y.; Goktas, M.; Tekinay, A. B.; Guler, M. O.; Hauser, C. A.; Mitraki, A., Self-Assembled Proteins and Peptides as Scaffolds for Tissue Regeneration. *Adv Healthc Mater* **2015,** *4* (16), 2557-86.

437. Tao, K.; Wang, J.; Zhou, P.; Wang, C.; Xu, H.; Zhao, X.; Lu, J. R., Self-assembly of short abeta(16-22) peptides: effect of terminal capping and the role of electrostatic interaction. *Langmuir* **2011,** *27* (6), 2723-30.

438. Tenidis, K.; Waldner, M.; Bernhagen, J.; Fischle, W.; Bergmann, M.; Weber, M.; Merkle, M. L.; Voelter, W.; Brunner, H.; Kapurniotu, A., Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties. *J Mol Biol* **2000,** *295* (4), 1055-71.

439. Kasai, S.; Ohga, Y.; Mochizuki, M.; Nishi, N.; Kadoya, Y.; Nomizu, M., Multifunctional peptide fibrils for biomedical materials. *Biopolymers* **2004,** *76* (1), 27-33.

440. Kasotakis, E.; Mossou, E.; Adler-Abramovich, L.; Mitchell, E. P.; Forsyth, V. T.; Gazit, E.; Mitraki, A., Design of metal-binding sites onto self-assembled peptide fibrils. *Biopolymers* **2009,** *92* (3), 164-72.

441. Terzaki, K.; Kalloudi, E.; Mossou, E.; Mitchell, E. P.; Forsyth, V. T.; Rosseeva, E.; Simon, P.; Vamvakaki, M.; Chatzinikolaidou, M.; Mitraki, A.; Farsari, M., Mineralized self-assembled peptides on 3D laser-made scaffolds: a new route toward 'scaffold on scaffold' hard tissue engineering. *Biofabrication* **2013,** *5* (4), 045002.

442. Li, D.; Furukawa, H.; Deng, H.; Liu, C.; Yaghi, O. M.; Eisenberg, D. S., Designed amyloid fibers as materials for selective carbon dioxide capture. *Proc Natl Acad Sci U S A* **2014,** *111* (1), 191-6.

443. Li, D.; Jones, E. M.; Sawaya, M. R.; Furukawa, H.; Luo, F.; Ivanova, M.; Sievers, S. A.; Wang, W.; Yaghi, O. M.; Liu, C.; Eisenberg, D. S., Structure-based design of functional amyloid materials. *J Am Chem Soc* **2014,** *136* (52), 18044-51.

444. Gazit, E., A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J* **2002,** *16* (1), 77-83.

445. Gazit, E., Reductionist Approach in Peptide-Based Nanotechnology. *Annu Rev Biochem* **2018,** *87*, 533-553.

446. Zhang, S.; Holmes, T.; Lockshin, C.; Rich, A., Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proc Natl Acad Sci U S A* **1993,** *90* (8), 3334-8.

447. Gelain, F.; Horii, A.; Zhang, S., Designer self-assembling peptide scaffolds for 3-d tissue cell cultures and regenerative medicine. *Macromol Biosci* **2007,** *7* (5), 544-51.

448. Yanlian, Y.; Ulung, K.; Xiumei, W.; Horii, A.; Yokoi, H.; Shuguang, Z., Designer self-assembling peptide nanomaterials. *Nano Today* **2009,** *4* (2), 193-210.

449. Kim, S.; Kim, J. H.; Lee, J. S.; Park, C. B., Beta-Sheet-Forming, Self-Assembled Peptide Nanomaterials towards Optical, Energy, and Healthcare Applications. *Small* **2015,** *11* (30), 3623-40.

450. Zhang, S.; Holmes, T. C.; DiPersio, C. M.; Hynes, R. O.; Su, X.; Rich, A., Self-complementary oligopeptide matrices support mammalian cell attachment. *Biomaterials* **1995,** *16* (18), 1385-93.

451. Yokoi, H.; Kinoshita, T.; Zhang, S., Dynamic reassembly of peptide RADA16 nanofiber scaffold. *Proc Natl Acad Sci U S A* **2005,** *102* (24), 8414-9.

452. Lampel, A.; Ulijn, R. V.; Tuttle, T., Guiding principles for peptide nanotechnology through directed discovery. *Chem Soc Rev* **2018,** *47* (10), 3737-3758.

453. Adler-Abramovich, L.; Aronov, D.; Beker, P.; Yevnin, M.; Stempler, S.; Buzhansky, L.; Rosenman, G.; Gazit, E., Self-assembled arrays of peptide nanotubes by vapour deposition. *Nat Nanotechnol* **2009,** *4* (12), 849-54.

454. Gorbitz, C. H., Nanotube formation by hydrophobic dipeptides. *Chemistry* **2001,** *7* (23), 5153-9.

455. Mondal, S.; Adler-Abramovich, L.; Lampel, A.; Bram, Y.; Lipstman, S.; Gazit, E., Formation of functional super-helical assemblies by constrained single heptad repeat. *Nat Commun* **2015,** *6*, 8615.

456. Reches, M.; Gazit, E., Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science* **2003,** *300* (5619), 625-7.

457. Ben-Zvi, O.; Grinberg, I.; Orr, A. A.; Noy, D.; Tamamis, P.; Yacoby, I.; Adler-Abramovich, L., Protection of Oxygen-Sensitive Enzymes by Peptide Hydrogel. *ACS Nano* **2021,** *15* (4), 6530-6539.

458. Jayawarna, V.; Ali, M.; Jowitt, T. A.; Miller, A. E.; Saiani, A.; Gough, J. E.; Ulijn, R. V., Nanostructured hydrogels for three-dimensional cell culture through self-assembly of fluorenylmethoxycarbonyl-dipeptides. *Adv Mater Weinheim* **2006,** *18* (5), 611-+.

459. Mahler, A.; Reches, M.; Rechter, M.; Cohen, S.; Gazit, E., Rigid, self-assembled hydrogel composed of a modified aromatic dipeptide. *Adv Mater Weinheim* **2006,** *18* (11), 1365-+.

460. Dudukovic, N. A.; Zukoski, C. F., Mechanical properties of self-assembled Fmoc-diphenylalanine molecular gels. *Langmuir* **2014,** *30* (15), 4493-500.

461. Yemini, M.; Reches, M.; Gazit, E.; Rishpon, J., Peptide nanotube-modified electrodes for enzyme-biosensor applications. *Anal Chem* **2005,** *77* (16), 5155-9.

462. Aviv, M.; Halperin-Sternfeld, M.; Grigoriants, I.; Buzhansky, L.; Mironi-Harpaz, I.; Seliktar, D.; Einav, S.; Nevo, Z.; Adler-Abramovich, L., Improving the Mechanical Rigidity of Hyaluronic Acid by Integration of a Supramolecular Peptide Matrix. *ACS Appl Mater Interfaces* **2018,** *10* (49), 41883-41891.

463. Halperin-Sternfeld, M.; Ghosh, M.; Sevostianov, R.; Grigoriants, I.; Adler-Abramovich, L., Molecular co-assembly as a strategy for synergistic improvement of the mechanical properties of hydrogels. *Chem Commun (Camb)* **2017,** *53* (69), 9586-9589.

464. Ryan, D. M.; Doran, T. M.; Nilsson, B. L., Stabilizing self-assembled Fmoc-F5-Phe hydrogels by co-assembly with PEG-functionalized monomers. *Chem Commun (Camb)* **2011,** *47* (1), 475-7.

465. Schnaider, L.; Ghosh, M.; Bychenko, D.; Grigoriants, I.; Ya'ari, S.; Shalev Antsel, T.; Matalon, S.; Sarig, R.; Brosh, T.; Pilo, R.; Gazit, E.; Adler-Abramovich, L., Enhanced Nanoassembly-Incorporated Antibacterial Composite Materials. *ACS Appl Mater Interfaces* **2019,** *11* (24), 21334-21342.

466. Ryan, D. M.; Anderson, S. B.; Nilsson, B. L., The influence of side-chain halogenation on the self-assembly and hydrogelation of Fmoc-phenylalanine derivatives. *Soft Matter* **2010,** *6* (14), 3220-3231.

467. Rajbhandary, A.; Brennessel, W. W.; Nilsson, B. L., Comparison of the Self-Assembly Behavior of Fmoc-Phenylalanine and Corresponding Peptoid Derivatives. *Cryst Growth Des* **2018,** *18* (2), 623-632.

468. Liyanage, W.; Nilsson, B. L., Substituent Effects on the Self-Assembly/Coassembly and Hydrogelation of Phenylalanine Derivatives. *Langmuir* **2016,** *32* (3), 787-799.

469. Beker, P.; Rosenman, G., Bioinspired nanostructural peptide materials for supercapacitor electrodes. *J Mater Res* **2010,** *25* (8), 1661-1666.

470. Bank-Srour, B.; Becker, P.; Krasovitsky, L.; Gladkikh, A.; Rosenberg, Y.; Barkay, Z.; Rosenman, G., Physical vapor deposition of peptide nanostructures. *Polym J* **2013,** *45* (5), 494-503.

471. Handelman, A.; Kuritz, N.; Natan, A.; Rosenman, G., Reconstructive Phase Transition in Ultrashort Peptide Nanostructures and Induced Visible Photoluminescence. *Langmuir* **2016,** *32* (12), 2847-62.

472. Nikitin, T.; Kopyl, S.; Shur, V. Y.; Kopelevich, Y. V.; Kholkin, A. L., Low-temperature photoluminescence in self-assembled diphenylalanine microtubes. *Physics Letters A* **2016,** *380* (18-19), 1658-1662.

473. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. 3rd ed.; Springer US: Boston, MA, 2006.

474. Vladimirov, Y. A., Primary steps of photochemical reactions in proteins and aromatic amino-acids: a review. *Photochem Photobiol* **1965,** *4* (3), 369-84.

475. Zuclich, J. A.; Maki, A. H., Protein triplet states. *Top Curr Chem* **1975,** (54), 115-63.

476. Lee, J. S.; Yoon, I.; Kim, J.; Ihee, H.; Kim, B.; Park, C. B., Self-assembly of semiconducting photoluminescent peptide nanowires in the vapor phase. *Angew Chem Int Ed Engl* **2011,** *50* (5), 1164-7.

477. Yan, X.; Su, Y.; Li, J.; Fruh, J.; Mohwald, H., Uniaxially oriented peptide crystals for active optical waveguiding. *Angew Chem Int Ed Engl* **2011,** *50* (47), 11186-91.

478. Amdursky, N.; Beker, P.; Koren, I.; Bank-Srour, B.; Mishina, E.; Semin, S.; Rasing, T.; Rosenberg, Y.; Barkay, Z.; Gazit, E.; Rosenman, G., Structural transition in peptide nanotubes. *Biomacromolecules* **2011,** *12* (4), 1349-54.

479. Manchineella, S.; Govindaraju, T., Molecular Self-Assembly of Cyclic Dipeptide Derivatives and Their Applications. *Chempluschem* **2017,** *82* (1), 88-106.

480. Choi, S. J.; Jeong, W. J.; Kang, S. K.; Lee, M.; Kim, E.; Ryu du, Y.; Lim, Y. B., Differential self-assembly behaviors of cyclic and linear peptides. *Biomacromolecules* **2012,** *13* (7), 1991-5.

481. Jeong, W. J.; Choi, S. J.; Choi, J. S.; Lim, Y. B., Chameleon-like self-assembling peptides for adaptable biorecognition nanohybrids. *ACS Nano* **2013,** *7* (8), 6850-7.

482. Tao, K.; Fan, Z.; Sun, L.; Makam, P.; Tian, Z.; Ruegsegger, M.; Shaham-Niv, S.; Hansford, D.; Aizen, R.; Pan, Z.; Galster, S.; Ma, J.; Yuan, F.; Si, M.; Qu, S.; Zhang, M.; Gazit, E.; Li, J., Quantum confined peptide assemblies with tunable visible to near-infrared spectral range. *Nat Commun* **2018,** *9* (1), 3217.

483. Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D., Structural chemistry of a green fluorescent protein Zn biosensor. *J Am Chem Soc* **2002,** *124* (14), 3522-4.

484. Zhou, L.; Li, S.; Su, Y.; Yi, X.; Zheng, A.; Deng, F., Interaction between histidine and Zn(II) metal ions over a wide pH as revealed by solid-state NMR spectroscopy and DFT calculations. *J Phys Chem B* **2013,** *117* (30), 8954-65.

485. Tamamis, P.; Kasotakis, E.; Archontis, G.; Mitraki, A., Combination of theoretical and experimental approaches for the design and study of fibril-forming peptides. *Methods Mol Biol* **2014,** *1216*, 53-70.

486. Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A., Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **2007,** *23* (19), 2625-7.

487. Seeber, M.; Felline, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F., Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J Comput Chem* **2011,** *32* (6), 1183-94.

488. Tamamis, P.; Adler-Abramovich, L.; Reches, M.; Marshall, K.; Sikorski, P.; Serpell, L.; Gazit, E.; Archontis, G., Self-assembly of phenylalanine oligopeptides: insights from experiments and simulations. *Biophys J* **2009,** *96* (12), 5020-9.

489. Facchetti, A., Semiconductors for organic transistors. *Materials Today* **2007,** *10* (3), 28-37.

490. Gunes, S.; Neugebauer, H.; Sariciftci, N. S., Conjugated polymer-based organic solar cells. *Chem Rev* **2007,** *107* (4), 1324-38.

491. Pansieri, J.; Josserand, V.; Lee, S. J.; Rongier, A.; Imbert, D.; Sallanon, M. M.; Kovari, E.; Dane, T. G.; Vendrely, C.; Chaix-Pluchery, O.; Guidetti, M.; Vollaire, J.; Fertin, A.; Usson, Y.; Rannou, P.; Coll, J. L.; Marquette, C.; Forge, V., Ultraviolet-visible-near-infrared optical properties of amyloid fibrils shed light on amyloidogenesis. *Nat Photonics* **2019,** *13* (7), 473-+.

492. Wang, C.; Dong, H.; Jiang, L.; Hu, W., Organic semiconductor crystals. *Chem Soc Rev* **2018,** *47* (2), 422-500.

493. Wang, Y.; Zhang, Z.; Xu, L.; Li, X.; Chen, H., Hydrogels of halogenated Fmoc-short peptides for potential application in tissue engineering. *Colloids Surf B Biointerfaces* **2013,** *104*, 163-8.

494. Ya'ari, S.; Halperin-Sternfeld, M.; Rosin, B.; Adler-Abramovich, L., Surface Modification by Nano-Structures Reduces Viable Bacterial Biofilm in Aerobic and Anaerobic Environments. *Int J Mol Sci* **2020,** *21* (19), 7370.

495. Cohen-Gerassi, D.; Arnon, Z. A.; Guterman, T.; Levin, A.; Ghosh, M.; Aviv, M.; Levy, D.; Knowles, T. P. J.; Shacham-Diamand, Y.; Adler-Abramovich, L., Phase Transition and Crystallization Kinetics of a Supramolecular System in a Microfluidic Platform. *Chem. Mater.* **2020,** *32* (19), 8342-8349.

496. Halperin-Sternfeld, M.; Ghosh, M.; Sevostianov, R.; Grigoriants, I.; Adler-Abramovich, L., Molecular co-assembly as a strategy for synergistic improvement of the mechanical properties of hydrogels. *Chem Commun (Camb)* **2017,** *53* (69), 9586-9589.

497. Makam, P.; Gazit, E., Minimalistic peptide supramolecular co-assembly: expanding the conformational space for nanotechnology. *Chem Soc Rev* **2018,** *47* (10), 3406-3420.

498. Zhang, H.; Wang, Y. J.; Xu, F., Impact of the subtle differences in MMP-12 structure on Glide-based molecular docking for pose prediction of inhibitors. *Journal of Molecular Structure* **2014,** *1076*, 153-159.

499. Prieto-Martínez, F. D.; Arciniega, M.; Medina-Franco, J. L., Acoplamiento Molecular: Avances Recientes y Retos. *TIP Revista Especializada en Ciencias Químico-Biológicas* **2018,** *21*.

500. Mohan, R. R.; Wilson, M.; Gorham, R. D., Jr.; Harrison, R. E. S.; Morikis, V. A.; Kieslich, C. A.; Orr, A. A.; Coley, A. V.; Tamamis, P.; Morikis, D., Virtual Screening of Chemical Compounds for Discovery of Complement C3 Ligands. *ACS Omega* **2018,** *3* (6), 6427-6438.

501. Sood, A. J.; Viner, C.; Hoffman, M. M., DNAmod: the DNA modification database. *J Cheminform* **2019,** *11* (1), 30.

502. Kumar, A.; Yoluk, O.; MacKerell, A. D., Jr., FFParam: Standalone package for CHARMM additive and Drude polarizable force field parametrization of small molecules. *J Comput Chem* **2020,** *41* (9), 958-970.

503. Guidance for Water Quality-based Decisions The TMDL Process: The TMDL Process. U.S. Environmental Protection Agency (EPA): Washington, D.C., 1991.

504. Monosson, E., Chemical mixtures: considering the evolution of toxicology and chemical assessment. *Environ Health Perspect* **2005,** *113* (4), 383-90.

505. Tamamis, P.; Kasotakis, E.; Mitraki, A.; Archontis, G., Amyloid-like self-assembly of peptide sequences from the adenovirus fiber shaft: insights from molecular dynamics simulations. *J Phys Chem B* **2009,** *113* (47), 15639-47.