DESIGN OF PEPTIDE INHIBITORS FOR INFLUENZA A NON-STRUCTURAL PROTEIN 1

by

PATRICK LEONARD NOSKER

A Dissertation submitted to the

Graduate School – New Brunswick and The Graduate School of Biomedical Sciences

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Programs in Biochemistry and Quantitative Biomedicine

Written under the direction of

Vikas Nanda

And

Gaetano Montelione

And approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October 2016

ABSTRACT OF THE DISSERTATION

DESIGN OF PEPTIDE INHIBITORS FOR INFLUENZA A NON-STRUCTURAL PROTEIN 1

By PATRICK LEONARD NOSKER

Dissertation Directors:

Vikas Nanda, Ph.D. and Gaetano Montelione, Ph.D.

Influenza A is responsible for tens of thousands of annual fatalities in the United States. The conventional medical strategy relies on prevention of infection using vaccines where are mildly effective at best and require yearly updates to reflect the expected prevalent strains, and anti-viral drugs for which there is significant clinical resistance. With this in mind, the non-structural protein 1 of influenza A (NS1A) is an appropriate target with high conservation and a multitude of anti-immune effects. It features an array of high surface area interactions which may be targetable using peptides.

We are developing molecular design tools to reliably create short (<20 amino acid) molecules with half-life and stability sufficient for drug-like function. In this thesis, I will explore D-amino acid motifs for increasing stability of peptides in the ideal Trp-cage peptide system and implement designed motifs in peptides designed to target the

NS1A protein. Finally, we will demonstrate inhibition of the NS1A protein using our designed peptides.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my advisors, Dr. Vikas Nanda and Dr. Gaetano T. Montelione, for being so supportive during my graduate career. Without their considerable guidance, support, and encouragement, the work presented in this dissertation would not have been possible.

I would also like to extend my sincere gratitude to the other members of my thesis committee, Dr. Ann Stock, Dr. Monica Roth, and Dr. David Case for their valuable suggestions and help. From my initial thesis proposal defense to my last committee meeting, my committee members have provided me deep insight into the methods of science, critical thinking, and have helped me hone my interdisciplinary thesis topic to one that is cohesive.

I would like to thank the members of Dr. Vikas Nanda's lab, former and present, who have helped me throughout my thesis. Dr. I. John Khan, Dr. Jim Stapleton, Dr. John Kim, Dr. Kenneth McGuinness, Dr. Daniel Hsieh, Dr. Stefan Senn, Dr. Hagai Hraanan, Dr. Fei Xu, Jose James, Fatu Conteh, Kaiser Loell and Daniel Grisham for our discussions and comradery. In addition, I am very thankful to Dr. Srinivas Annavarapu, Dr. Agustina Rodriguez-Granillo, Nida Hasan, and Sumana Giddu for their help with the D-amino acid Trp-cage project. I would like to specifically thank Dr. Avanish Parmar for his help with learning circular dichroism, dynamic light scattering, and electron microscopy in addition to being a great colleague and inviting me to give my first international seminar talk at the University of Hyderabad. I would also like to thank Douglas Pike for his assistance with learning molecular dynamics and ProtCAD, great

discussions, and help with the influenza A NS1 RNA-binding domain project as well as turning ProtEvolver from a basic, slow, and inaccurate tool to one that has the ability to efficiently design candidate peptides for binding inhibition. My Nanda labmates have made my time in the lab a fruitful one with many new friendships forged.

I would also like to thank the members of Dr. Gaetano Montelione's lab, former and present, who have helped me throughout my thesis. Dr. Li Ma, Dr. Jim Aramini, Dr. G.V.T. Swapna, Keith Hamilton, Amy Suhotliv, and Emily Grasso have helped me learn NMR, efficient protein expression and labeling techniques, and have contributed deeply to my understanding of structural biology and the influenza virus. In addition, Dr. Li Ma, Dr. Jim Aramini, Dr. G.V.T. Swapna and Keith Hamilton have contributed significantly to my study of the NS1 effector domain studies. Dr. Jim Aramini and Dr. G.V.T. Swapna have assited me greatly on the D-amino acid Trp-cage project. Dr. Li-Chung Ma, Dr. Jim Aramini, and Emily Grasso have contributed to the NS1 RBD project. My Montelione labmates have helped me develop as a structural biologist and I greatly appreciate their help throughtout the past several years.

In addition to my labmates, I would like to thank my collaborators and friends at Rutgers who have helped me complete my thesis. I would like to extend my thanks to Dr. John Tainer and Dr. Jeff Perry at The Scripps Research Institute for introducing me to small angle x-ray scattering. In addition, I am appreciate of Dr. Ross Walker from the University of California, San Diego for helping me with CUDA-accelerated molecular dynamics. I am thankful for the opportunity to learn about toxins by collaborating with Dr. Richard Hann and Dr. Wanda Velez at the Universidad Central de Caribe in Bayamon, Puerto Rico. Finally, I am thankful for my CABM building-mates for their

advice and help throughout the years. Dr. Abdul Khan, Dr. Christopher Barbieri, and Dr. Paul Leonard have been helpful with teaching me the methods of isothermal calorimetry. I am thankful for Dr. Samantha Yost and Dr. Paul R. Buckland's help in completing this written thesis.

Finally, I am thankful for those that have provided funding for me to complete my thesis. I am very thankful for the NIH for supporting me on a fellowship for my first year, and as a graduate assistant on Dr. Vikas Nanda's R01-GM089949 for my second year. My third and fourth years were funded thanks to the NIH and Dr. Gary Brewer's training grant T32-AI7403, and finally my last year thanks to the US Department of Education and the Center for Integrative Proteomics Research's GAANN fellowship. These sources of funding helped provide me the ability to work on my thesis.

DEDICATION

To my mother,

Dr. Ching Yi Chang and

My father,

Dr. Thomas Jerome Nosker

To my younger brother,

Kyle Thomas Nosker

To my beloved girlfriend,

Kristin Marie Blacklock

To my friends,

Jerimiah Booream,

AJ Jursik,

Eric Fox,

David Boehm, and

Michael Convente

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

AMBER	Assisted Model Building with Energy Refinement
CD	Circular Dichroism
(E)COSY	(Exclusive) Correlation Spectroscopy
ED	Effector Domain
FDA	Food and Drug Administration
FMOC	Fluorenylmethyloxycarbonyl chloride
FPLC	Fast Protein Liquid Chromatography
GCC	Gnu Compiler Collection
GLP-1	Glucagon-like peptide-1
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear single quantum coherence spectroscopy
kDa	Kilodalton (mass)
MRE	Mean Residue Ellipicity
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NS1A	Influenza A Non-Structural Protein 1
ProtCAD	Protein Computer Aided Design
RBD	RNA-Binding Domain
RMSD	Root-mean-square deviation
TOCSY	Total Correlation Spectroscopy
WT	Wild-type

INTRODUCTION

Protein design and engineering

There are two common approaches currently used to design proteins: *de novo* design involves building a new protein from scratch, while protein redesign uses known structure and sequence data as a starting framework to build a new protein. Both approaches most often utilize physics-based calculations, but some also use knowledge-based potentials in order to help guide the design of a novel protein. Physics-based design approaches often consider bonds between atoms in a simplified Newtonian mechanics scheme where bonds are modeled as springs using simple physics and statistical potentials. Knowledge-based design schemes are usually based upon previously demonstrated structural information such as that located in the Protein Data Bank (PDB). In this scheme, statistical data from known structures are used to guide design. Validation and testing of the newly designed proteins is often done using structural and biophysical approaches such as X-ray crystallography or nuclear magnetic resonance (NMR) on synthesized or recombinant expressed proteins. For designed proteins with an intended activity, various assays may be used to examine whether the intended activity exists in the design.

Over the past several decades, the field of protein design and engineering has evolved, integrating molecular force fields that approximate reality and basic libraries of known amino acid conformations. This progress has led to the creation of sophisticated protein design tools such as ROSETTA, which incorporates a physics-based force field

derived from CHARMM, as well as knowledge-based functions incorporating amino acid propensities and rotamer probabilities.[1–3]

A major problem constraining the field of protein design is the difficulty in predicting the fold of an amino acid primary sequence. Because finding the lowest-energy state of a protein is classified as a problem with the true and ultimate solution being impossible to solve in polynomial time (referred to as NP-Hard), careful approximations and simplifications are used in order to find a solution that samples large numbers of solutions rapidly.[4] Some of these approaches include dead-end elimination, branch-and-bound algorithms, and Monte Carlo simulation. Dead-end elimination iteratively searches and removes amino acid rotamers that are unacceptable when looking for a global energy minima, such as those with obvious steric clashes.[5] Branch-and-bound algorithms leverage the A* search algorithm and use heuristics to preserve promising fold-intermediates and further branching of partial solutions to derive the lowest energy solution.[6] Monte Carlo simulations use a probability distribution of possible results to iteratively calculate changes to the input information until an approximate solution is found. Several tools, including Rosetta, rely on a Monte Carlo approach. This method relies on brute-force random computational mutagenesis while lower energy sequences are saved and sequences with higher energy are accepted at a rate corresponding to a Boltzmann probability (the probability that a certain state may exist) of a given acceptable system temperature (allowance of energetic violations).[7]

Once a designed protein is synthesized, experimental approaches are used to validate the structure and function. If the protein design was effective, the predicted structure will be similar to the true 3D structure of the folded protein. Even if the

structure is correct, in the case of enzyme design, the designed protein still may not perform as intended. Further validation of activity must be completed. Likewise, the design of a protein with a target partner could require experimental approaches to measure interaction with its intended target. Various projects including the Critical Assessment of protein Structure Prediction (CASP) exist to help refine protein structure prediction software by validating predicted structures with experimentally determined structures.[8] This data is then used to further optimize the structure prediction software.[9]

In this project we use protein design, molecular dynamics, and biophysical evaluation methods, including NMR, to evaluate the effects of amino acid stereochemistry on protein structure with the ultimate goal of applying new techniques to stabilize potentially therapeutic peptides. In the following introductory sections, we introduce relevant concepts in stereochemistry, motifs used to cap α-helices, methods in molecular dynamics and NMR, as well as our therapeutic target of influenza.

Amino acid chirality in proteins

As this thesis examines specific effects of amino acid stereochemistry on protein structure and stability, we briefly review some fundamental principles regarding amino acid stereochemistry. Proteins found throughout living organisms are made up of amino acid building blocks which feature a free amine (N-terminus) and carboxylic acid (C-terminus) as well as a varying sidechain group which are covalently bonded directionally from N to C-terminus in which new amino acids are added to the carbonyl end in order to create a protein polypeptide chain through the process of ribosomal translation. The

sidechain group of the amino acid can vary with 20 or more varieties and can feature vastly different chemistry. The chemical differences in the amino acid sidechain can exhibit positive (+) or negative (-) charge, polar groups such as hydroxyl (-OH), hydrophobicity, or aromaticity.

Almost all amino acids observed in natural proteins possess *Levo*-stereoisomerism (L-) except in rare circumstances (e.g. bacterial envelopes, antibacterial secretions) (Figure 0-1 - L vs D-).[10] D-amino acids have been found in older human tissues and are observed at an unusually high rate in neurodegenerative disease tissue.[11] In many cases, the original L-amino acid may undergo spontaneous racemization. Amyloid-related diseases such as Alzheimer's and Parkinson's diseases have been connected to D-amino acid racemization with D-serine observed in amyloids, while D-aspartate has been observed in the crystallin family proteins in cataract lenses.[12]

Glycine features a sidechain composed of only a hydrogen atom. Since the atom is the same as the already present hydrogen bound to the α-carbon as in other amino acids, it can exhibit structural characteristics consistent with both L- and D-amino acids by allowing low energy dihedral angles in structural space that are unfavorable for L-amino acids. For all other amino acids besides isoleucine and threonine, the L- and D-amino acid forms are equivalently mirrored with atom positions in the reflected positions from the alternate form. This effect is due to the presence of only a single chiral center, the α-carbon.

Isoleucine and threonine contain a second chiral center at the β-carbon position and therefore can exist as diastereomers.[13] The mirror of L-isoleucine and L-threonine produce D -isoleucine and D –threonine; however, two other enantiomeric identities

exist, D-allo-isoleucine and D-allo-threonine, in which the atoms bound to the β-carbon are switched (Figure 1-3 - Substituted amino acids in 10 position). Because of this, it is possible for four alternate amino acids to exist which have the same atomic contents: L-allo-isoleucine, L-allo-threonine, D-isoleucine, and D-threonine. We will use D-amino acids to build α-helix caps in order to impart increased stability of our designed peptides.

α-helix caps

If small peptides are to become a more feasible option for therapeutics, it is necessary to increase structural stability to prevent proteolytic degradation and reduce the likeliness for aggregation. Since our intended peptides are smaller than currently demonstrated soluble peptides with stable secondary structure, it is important to understand the current techniques used to stabilize short α-helices. Proteins are made up of a chain of amino acids (primary structure) which, due to chemical and physical effects, usually organize into structural elements (secondary, tertiary, quaternary structure). The secondary structural elements observed in natural proteins include α-helices and β-sheets in which hydrogen bonding between the amino acid backbone amide and carbonyl groups hold the protein in a limited set of conformations. In addition to α-helices and β-sheets, loops, turns, β-strands, bends, coils, and other elements exist in proteins. The hydrogen atom of the backbone amide group acts as a hydrogen bond donor, sharing its hydrogen with the oxygen of the backbone carbonyl group with usual observed distances of 2.2-4.0 Å and energetic contributions of 4-40 kcal/mol.[14]

Various motifs that serve as energetically favorable caps for α-helices have been identified for both N and C-termini. At the N-terminus, the sidechain of the first amino acid in position N can often hydrogen bond to the N+3 position backbone. Amino acids observed at high frequency in position N include threonine, serine, and asparagine, and often interact with amino acids glutamate and glutamine (Figure 0-3 - N-cap box motif). Likewise, the C-terminus often features one of two observed motifs: the Schellman or the α_L-motif.[15] The Schellman motif exhibits a pair of hydrogen bonds from the C amide hydrogen to C+5 carbonyl oxygen as well as the C+1 carbonyl oxygen to the C+4 amide hydrogen.[16,17] The α_L-motif features a glycine at the C-terminal position where it can take advantage of its innate D-amino acid characteristics and allow the backbone dihedral angle to reside in conformational space normally not available to most other amino acids. In this C-capping motif, an additional backbone hydrogen bond is gained. In addition to glycine, asparagine is able to situationally adopt D-configurations using its sidechain as a backbone mimic.

Recent advances in protein engineering have provided additional options for stabilizing α-helices. D-amino acids have been used as caps, demonstrating considerable enhancement of structural and thermal stability. By substituting D-alanine or D-glutamine for glycine in α_L cap regions where the glycine has a positive φ angle, the reduction of entropic space due to the bulky sidechain near the helix cap provides significant advantages in stability (

Figure 0-4 - D-glutamine capped in Trp-cage).[18–20] Addition of a D-amino acid restricts the allowed spatial arrangement to significantly less because L-conformation space is not favorably accessible. Likewise, adding a D-asparagine to the N-terminus can provide a positive energetic contribution to helix stability.[21] In order to examine the impact of D-

amino acids as caps, we will utilize molecular dynamics simulations to help optimize our α-helix cap designs.

Molecular dynamics simulations

Molecular dynamics (MD) is a computational tool that enables studying the physical motion of the atoms that comprise a system. In terms of biological MD, it is often used to study proteins and nucleotides. Various approaches exist in which degrees of realism can differ significantly. Coarse-grained approaches and other reduced representations of the system of study allow for faster simulation at the expense of detail and accuracy.[22] In this approach, pseudo-atoms represent groups or even entire amino acids (

Figure 0-2 - Representation of atomistic, hybrid, and coarse-grained molecular dynamics). For simulations of extremely large systems, this is often the only practical approach. Atomistic approaches often lie within one of three methods: *in vacuo,* implicit solvation, or explicit solvation simulations.[23] The *in vacuo* approach simulates behavior without environmental effects. Due to recent advances in computational power, *in vacuo* approaches are rarely used. Implicit solvation (also known as continuum solvation) simulates only the molecule of interest with a continuous solvent where solvent effects are represented by the average characteristics of the solvent used in simulation.[24,25] Explicit solvation simulations include atomistic representations of water molecules in the system, each with the ability to move and interact with other water molecules and the

molecule of interest. In terms of realism, the explicit solvent simulation usually exceeds the implicit simulation but is significantly slower and requires more computing power.

In all forms of MD, physics-based force fields are used to calculate the energetics of the system. Atoms are assigned initial velocities and positions with timestep frames used to update the representation.[26] Control of temperature, pressure, and volume are often used along with a thermal control system which allows energetically unfavorable events to occur occasionally.

With the recent advances in computing, MD simulations have become efficient enough to fully fold small proteins, such as the 20 amino acid Trp-cage mini-protein.[27] With massive parallelism using graphics processors, simulation speeds have increased by orders of magnitude in recent years.[28]

Simulations are also used to calculate various energies of the system including binding energy between two partners and energy of solvation, allowing rapid *in silico* screening of designed proteins and peptides.[29,30] While mathematical and physical assumptions exist in simulations, results derived from them often correlate with experimental results. In addition, the accuracy of MD simulations is sufficient to allow further refinement of structures generated from X-ray crystallography or NMR experiments, when given the appropriate restraints.[31] We will use MD simulations in order to stabilize designed peptides in the hopes of generating therapeutically viable biologics.

Peptide therapeutics

Biologics are a rapidly growing field in pharmaceutical science.[32] These therapeutic proteins, often antibodies, are significantly larger than the traditional small molecule compound. The increased size provides a higher interaction specificity than small drug-like molecules due to larger contact area with the target. Additionally, the customization of these biologics is simpler, using genetic approaches rather than chemical.[33] With these benefits come negative consequences: cost is often greater than that of a small molecule compound, the biologic is vulnerable to standard protein degradation pathways with proteolysis causing reduced half-life, and large proteins will be occluded from cell entry, restricting target molecules to extracellular spaces or on the cell surface. Occasionally, immunogenic responses are also observed due to an apparent systemic invasion from a foreign biological molecule.

Peptides are often smaller than other proteins and have molecular weights between that of small molecule compounds and the traditional antibody conjugate that makes up most biologic therapeutics. While still possessing the ability to create a large surface area contact with the target molecule, cell-penetrating peptides have been designed so they may enter the cell and provide activity internal activity akin to how many small molecule drugs function.[34] Still, the factors of immunogenicity and degradation exist. Studies of the immunogenicity of protein and peptide therapeutics have shown that anti-therapeutic antibodies are found in patients at levels from 0-44 of patients%, depending on the therapeutic.[35] Unfortunately, because of this wide range, it is difficult to predict before administration; however, this attribute has been exploited by using peptides as vaccine epitopes. New approaches have been derived to reduce the impact of both immunogenicity and degradation.

In order to be a viable therapeutic option, peptides must overcome the issue of stability.[36,37] As peptides are made up of peptide bonds as proteins are, peptide bond-cleaving proteases pose a significant threat to peptide therapeutics. Additionally, peptides are often more accessible to proteases due to low structural stability. It is predicted that by enhancing structural stability, solubility is, in turn, increased and the ability for proteases to access and cleave the peptide will be reduced.

Several peptide therapeutics are currently approved for the treatment disease. Insulin is one of the most widely used peptide drugs in existence with well-engineered recombinant production platforms. As insulin is a naturally occurring hormone, it is well suited to be delivered to patients with defective insulin regulation systems. Likewise, the exendin-4 and glucagon-like-peptide-1 peptides have been used to treat diabetes.[38] A peptide therapeutic has been shown to cause weight loss and improve insulin resistance in obese monkeys as well.[39] Peptide therapeutics can also be a viable option for the treatment of infectious diseases. D-amino acid-containing HIV entry inhibitor peptides have been discovered and are resistant to protease activity while conferring a highly potent binding mode.[40]

The large potential contact area of peptides to their target makes them well suited for inhibition of protein-protein interactions and protein-nucleotide interactions. This set of interaction targets has been considered the “high-hanging fruit” in drug discovery since most small molecules are too small to sufficiently block these interactions directly and often rely on allosteric changes for inhibition.[33] The methods of peptide cyclization or stapling may allow faster development of peptide therapeutics to reach these “fruit”

while increasing structural stability and potentially increasing their ability to enter cells.[32,41,42]

Here, we explore the use of D-amino acids to stabilize active forms of peptides as an alternative strategy for improving therapeutic efficacy. Our designed peptides will be built to inhibit pathways in the influenza virus in the hopes of furthering peptide-based anti-viral therapies.

Influenza virus

The influenza virus is a disease-causing agent responsible for tens of thousands of deaths per year in the United States alone.[43] Significant epidemic and pandemic outbreaks of influenza have occurred with the 1918 influenza pandemic causing an estimated 50-100 million deaths.[44] Vaccination programs exist but at best provide moderate protection overall with very little protection in older adults.[45] The low genomic stability of influenza as well as a wide variety of subtypes creates difficulties in long-term immunization and the predicted epidemiology of future strains is often incorrect causing resources to be wasted on an ineffective vaccine. Anti-viral therapies exist for post-infection treatment but many are useless against newer mutant strains.[46,47]

Influenza is comprised of two or three types: A, B, and occasionally C based upon the immunological relationship of the nucleocapsid proteins.[48] Influenza type A is the most studied and has the widest range of host infection. Influenza virus features a segmented, negative-sense RNA genome that is encapsulated by a host-derived lipid envelope. Viral hemagglutinin (HA) and neuraminidase (NA) glycoproteins, ion channel

(M2), and matrix protein (M1) are embedded in the membrane (Figure 0-5 - Schematic of the Influenza A viral particle). The surface proteins HA and NA trigger an antigenic immune response after host invasion. Inside the envelope are polymerase proteins (PA, PB1, and PB2) as well as non-structural proteins (NS1 and NS2). Each protein is encoded by an RNA segment while M1 and M2 are encoded by RNA segment 7; NS1 and NS2 are encoded by RNA segment 8.

The mechanism for influenza entry is receptor-mediated endocytosis. The viral particle is internalized by the cell after binding a cell surface receptor, after which the host-viral membrane fusion event is triggered by low pH in the endosome.[49] The viral RNAs enter the nucleus and are transcribed to positive-sense, complementary strands used for viral protein synthesis. Newly replicated negative sense RNAs are exported to progeny virions.[50,51] The viral mRNA synthesis requires hijacking host-cell RNA fragments in order to produce translatable viral mRNAs.[52] While viral protein synthesis is occurring, host-cell synthesis is significantly reduced, turning the host-cell into an efficient virus-producing factory and limiting the cell's ability to move into an anti-viral state by preventing the production of immune response elements.

Structure and biology of the NS1 protein from influenza A virus

NS1 and NS2 of influenza A virus are both encoded on the same RNA segment with a significant coding region overlap.[53] NS1 and NS2 have offset reading frames from each other so nucleotide mutations in the segment have a significant potential to affect both proteins. This region shows high sequence conservation due to this effect. Evidence

of this is observed with a single nucleotide mutation which causes a single amino acid change and causes temperature sensitivity of the virus.[54]

NS1 has many functions including sequestration of host mRNAs containing a poly(A) tail, inhibition of pre-mRNA splicing, and inhibition of the innate host immune response by binding double-stranded RNA (dsRNA), a pathogen-associated molecular pattern which typically triggers the innate immune response. Additionally, NS1 sequesters the host cleavage and polyadenylation specificity factor 30 protein (CPSF30) which is required for interferon maturation.[55] In the context of host interferon β (IFN β), NS1 has been shown to both pre-transcriptionally and post-transcriptionally limit IFN β production.

The NS1 protein is composed of two key domains made of 237 amino acids. The N-terminal domain has been shown to bind dsRNA and is referred to as the NTD or RNA-Binding Domain (RBD), ranging from amino acids 1-73. The NS1 C-terminal (CTD) or Effector Domain (ED) from amino acids 85-237 binds several partners, including itself as a homodimer, over the course of infection. The two domains are connected by a short flexible linker.

The NS1 RBD is composed of a tightly bound homodimer with a deep cleft (Figure 0-6 - Influenza A NS1 RNA binding domain). This cleft is positively charged due to several arginine and lysine amino acids. Double-stranded RNA has been shown to bind in a non-specific manner through an interaction between the arginine 37 sidechain and the polyphosphate backbone of the nucleotide.[56,57] Since influenza is a negative sense RNA virus which requires an intermediate double-stranded state during replication, it is suspected that NS1 binds dsRNA to prevent detection of viral dsRNA by host factors. In

humans, dsRNA is a strong trigger for the serine/threonine protein kinase R (PKR) and 2'-5'-oligoadenylate synthetase (OAS) pathways which trigger immune response function and host apoptotic response.[58] By preventing the cell machinery from detecting the dsRNA, the virus can continue replicating and maturing until the virus is ready to trigger apoptosis and viral release.

The ED of NS1 has been shown to bind to many partners. Crystallization studies have shown this domain as a homodimer in multiple configurations (Figure 0-7 - Influenza A NS1 effector domain homodimer). NMR experiments demonstrate binding at the same interface as seen in CPSF30 binding (Figure 0-8 - Influenza A NS1 effector domain/CPSF30 complex).[59–61] Other partners observed include the p53 apoptosis protein, nucleolin, ubiquitin, interleukin-6 receptor, and others. [62–64]

Recent models have predicted that the NS1 protein is a multimer with itself and dsRNA. In this model, the NS1 RBD homodimer binds dsRNA while the ED binds another dimer complex creating a large, dynamic assembly of NS1 proteins wrapped around the core dsRNA molecule. In this model, NS1 acts as a sheath for the RNA, completely encapsulating it from the host viral-detection systems.[65,66] We will use fluorine NMR to examine the structure and biology of the Influenza A NS1 protein.

Fluorine NMR in biology

Nuclear magnetic resonance techniques have provided significant insight into the structure and dynamics of biomolecules. Traditional ^{1}H, ^{13}C, and ^{15}N based techniques can provide deep insight on the connectivity and relative positioning of the atoms that

make up a protein; however, the abundance of great information often makes it difficult to assign resonance peaks and gather information without cumbersome mutagenesis and changing of experimental design.[67] A 30 kDa protein may contain hundreds of hydrogen atoms making 1 dimensional (1D) NMR experiments impossible to interpret. Even two or three dimensional experiments such as HSQC or NOESY experiments require tedious work for assignment, and often have poor signal-to-noise ratio for proteins larger than 20 kDa.

Fluorine NMR provides a tool that can be used to probe very specific dynamic effects in proteins. By strategically replacing a given amino acid with one featuring a single ^{19}F atom, such as 5-flouro-tryptophan, it is possible to greatly simplify the NMR spectra for a large protein, such that only a handful of resonance peaks are present. In addition to the ability to reduce the data set to a minimum, fluorine offers the feature of high sensitivity, no background signal unless fluorine is present in the working buffer, and a wide range of chemical shifts (spanning ~400 ppm vs ~15 ppm for ^{1}H) allowing for exquisite chemical shift sensitivity to environment and dynamics, easy assignment by site-directed mutation, and simpler data analysis.[68] Because of these attributes, scientists have used ^{19}F NMR for screening of small molecules for target proteins with well understood fluorine-labeled amino acid protein synthesis approaches.[69]

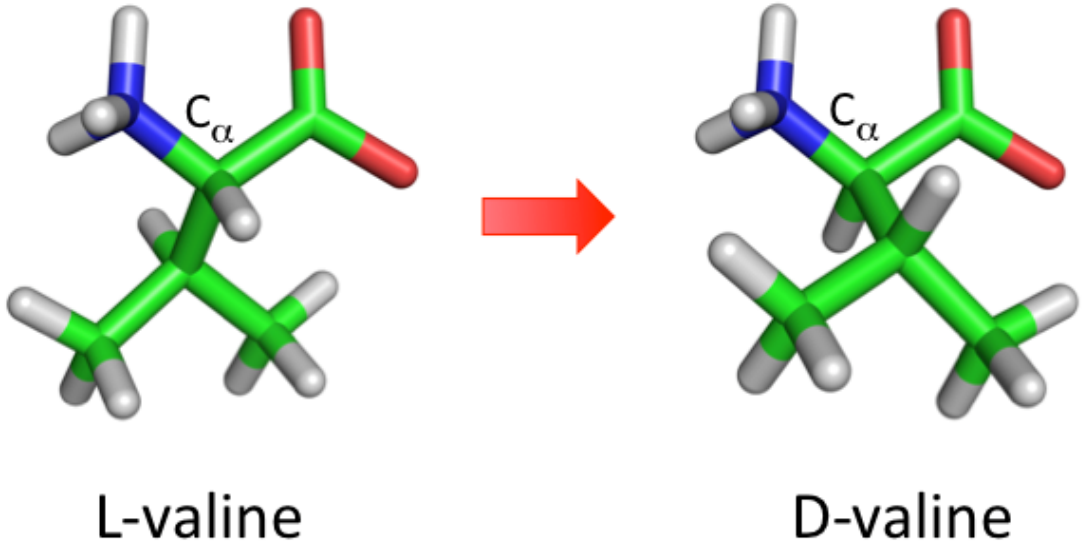


Figure 0-1 - L vs D-valine

The sidechain β carbon and α hydrogen are in opposite positions.

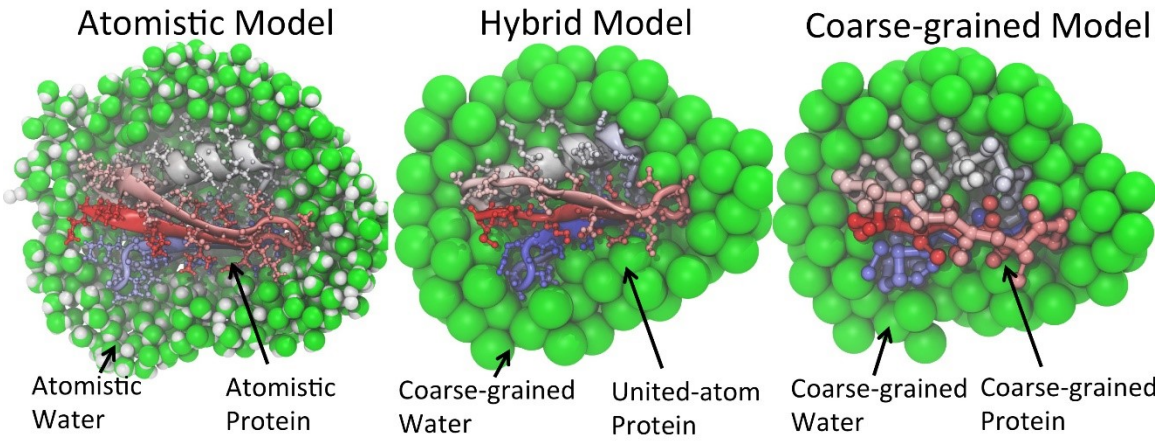


Figure 0-2 - Representation of atomistic, hybrid, and coarse-grained molecular dynamics

A representation of atomistic, hybrid, and coarse-grained molecular dynamics structures from a simulation of PDB 3GB1 as published on the University of Illinois – Urbana Champaign's Theoretical and Computational Biophysics Group website.[70]

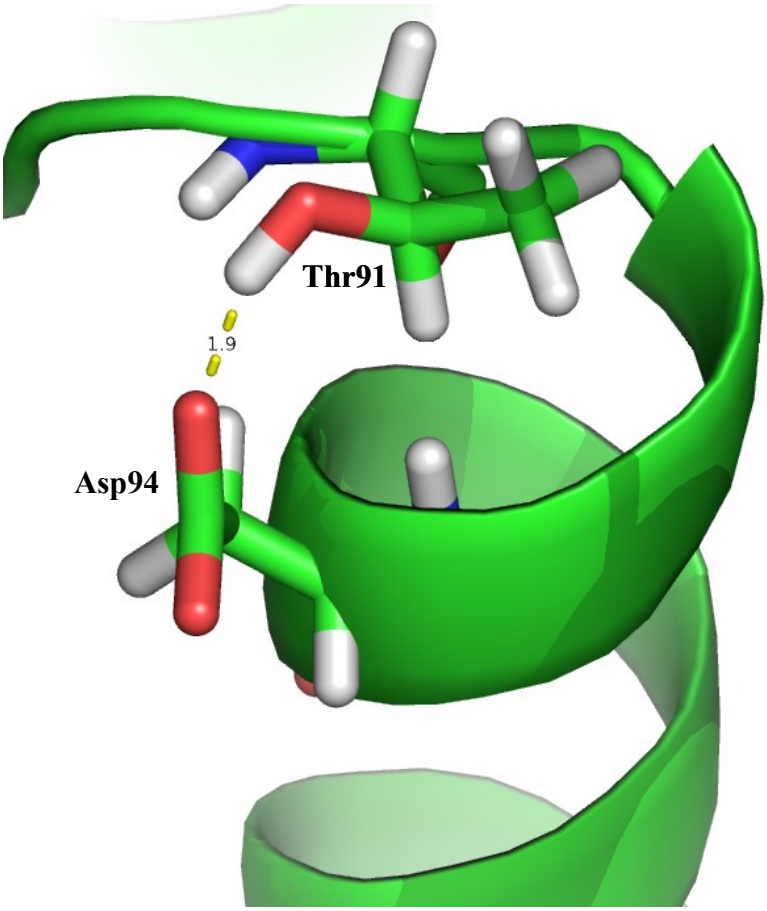


Figure 0-3 - N-cap box motif

The prototypical Serine/Threonine-X-X-Glutamate/Glutamine box cap is observed in the Apoferritin structure (PDB 1HRS).

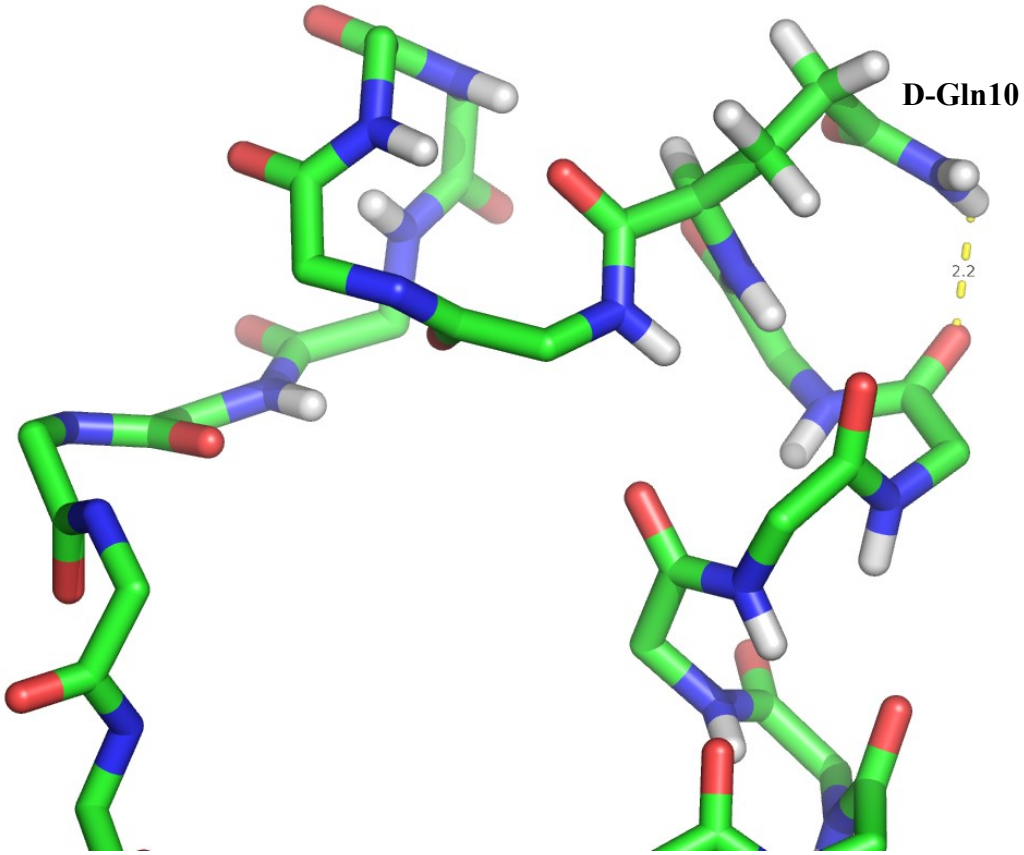


Figure 0-4 - D-glutamine capped in Trp-cage

C-terminal cap engineered to feature D-glutamine allowing for an additional hydrogen bond to form and increasing thermal stability (PDB 2LDJ).

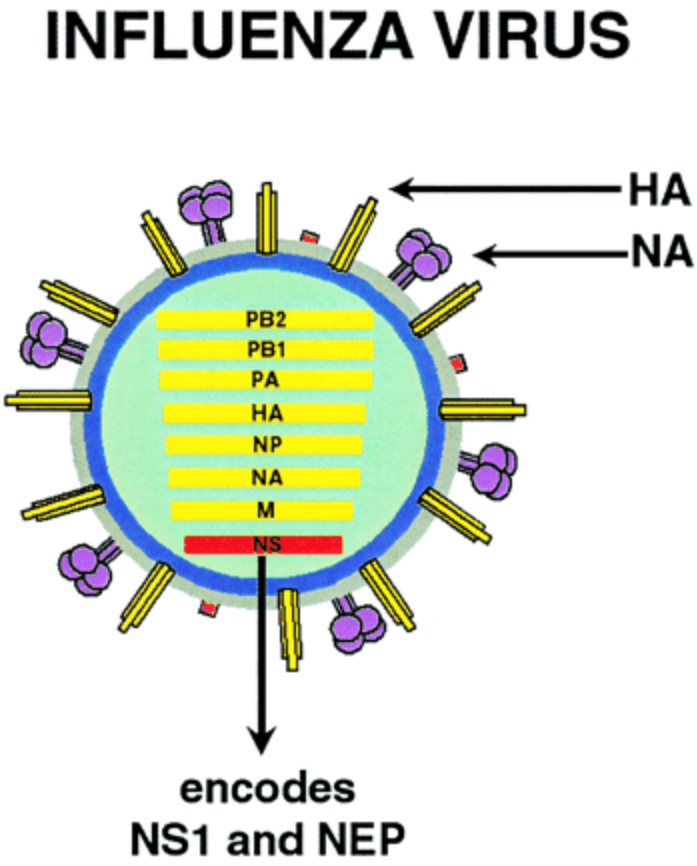


Figure 0-5 - Schematic of the Influenza A viral particle

The influenza A viral particle is composed of a coat comprised of HA and NA glycoproteins along with the M1 and M2 matrix proteins. Inside the particle are the coding RNA molecules PB1, PB2, PA, HA, NP, NA, M, and NS which code for the proteins that are necessary for influenza A replication.[71]

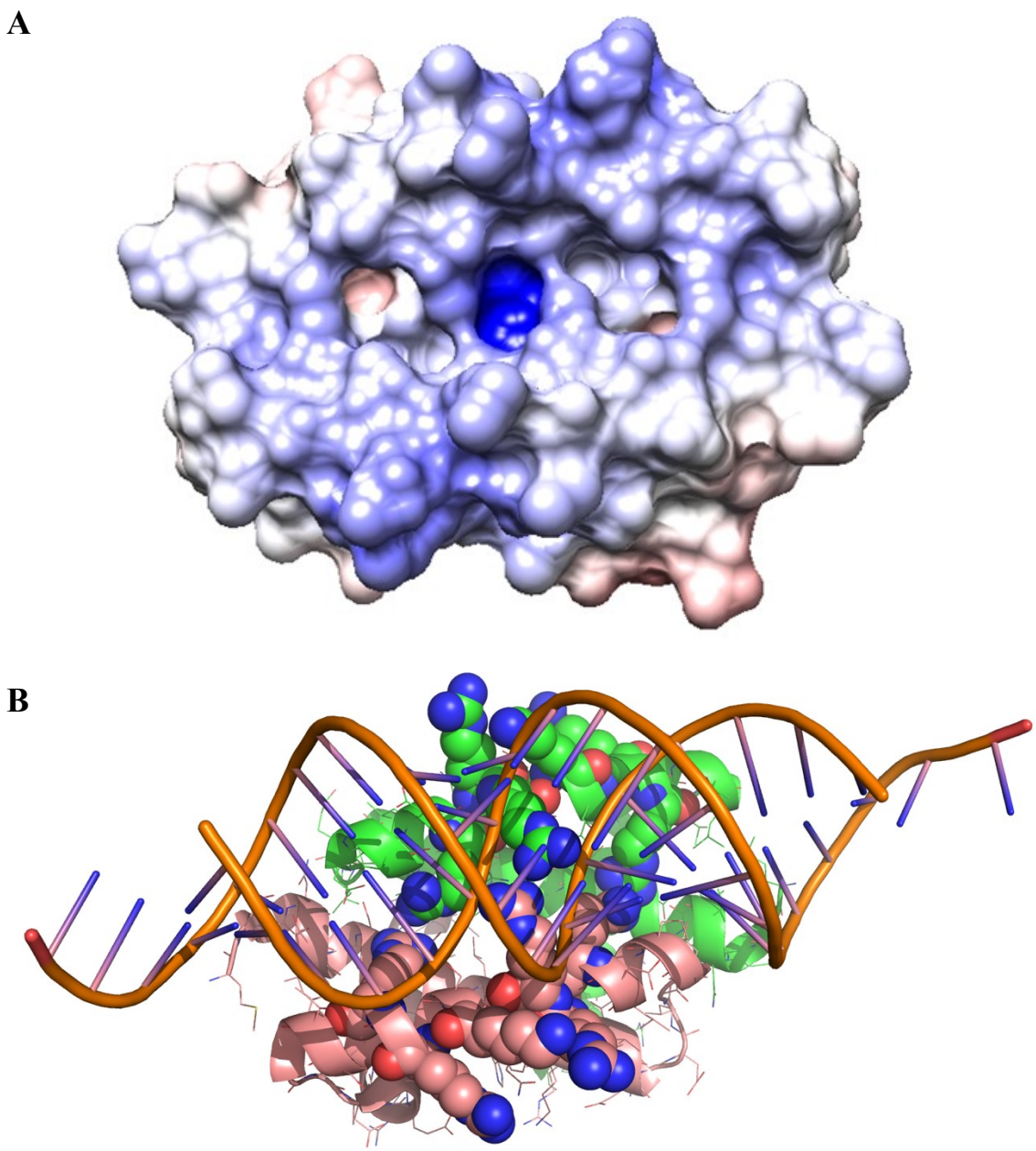


Figure 0-6 - Influenza A NS1 RNA binding domain

The NS1 RBD is a heterodimer composed of two monomer subunits. A. The RNA binding cleft is positively charged (PDB: 1NS1). B. A high concentration of arginine and lysine amino acids at the binding site present a strong interaction site with double-stranded RNA (PDB: 2ZKO). Arginine and lysine residues near the binding site are shown using a spherical representation.

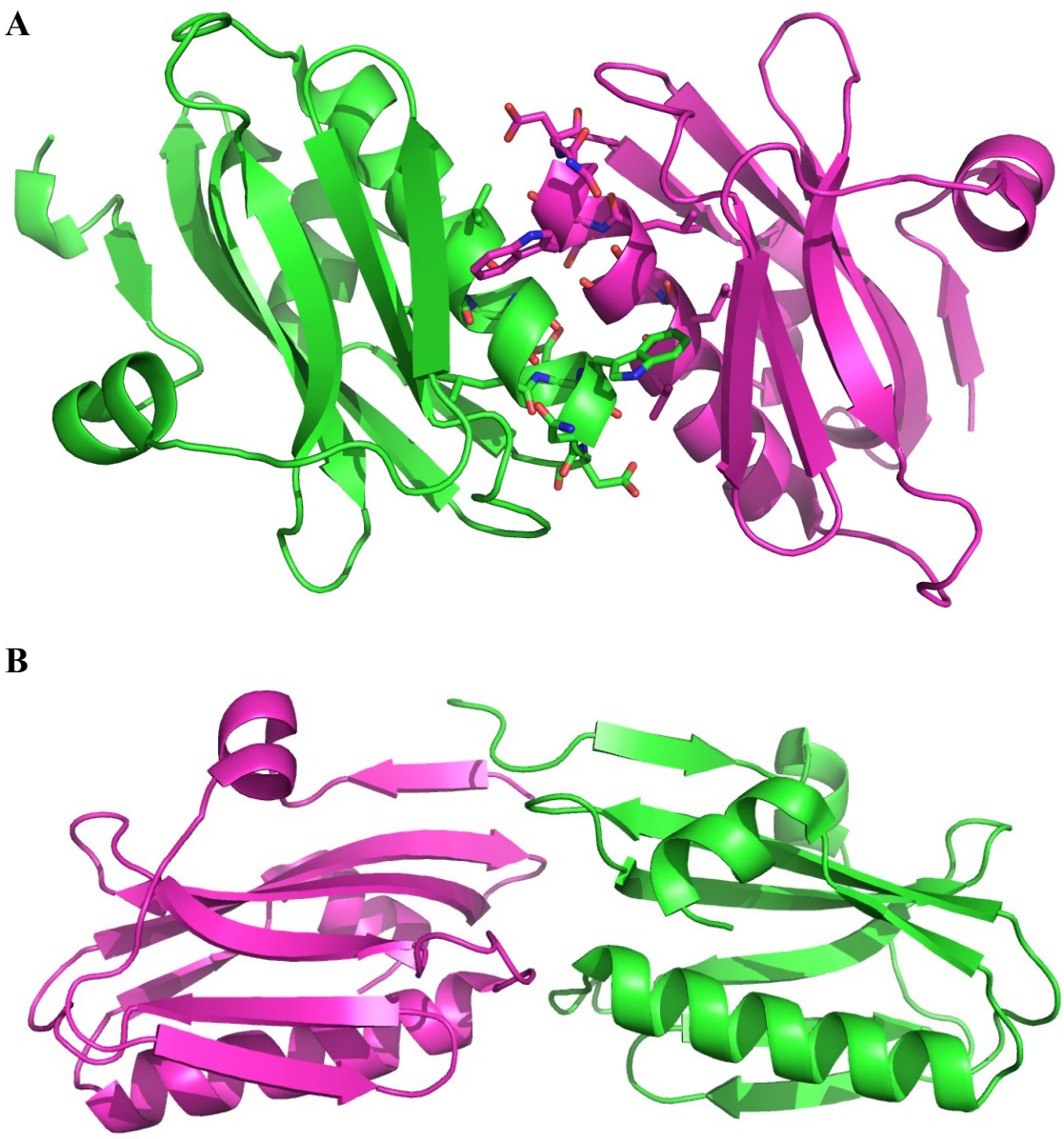


Figure 0-7 - Influenza A NS1 effector domain homodimer

The NS1 Effector Domain can form a homodimer with multiple dimer conformations observed in crystallization conditions. It is predicted that the preferred dimer form is that as observed in A (PDB: 3EE9) with NMR data consistent with this form. The tryptophan amino acids shown in sticks at the binding interface also play an important role in CPSF30 binding. B. Another observed dimer interface (PDB: 3O9T).

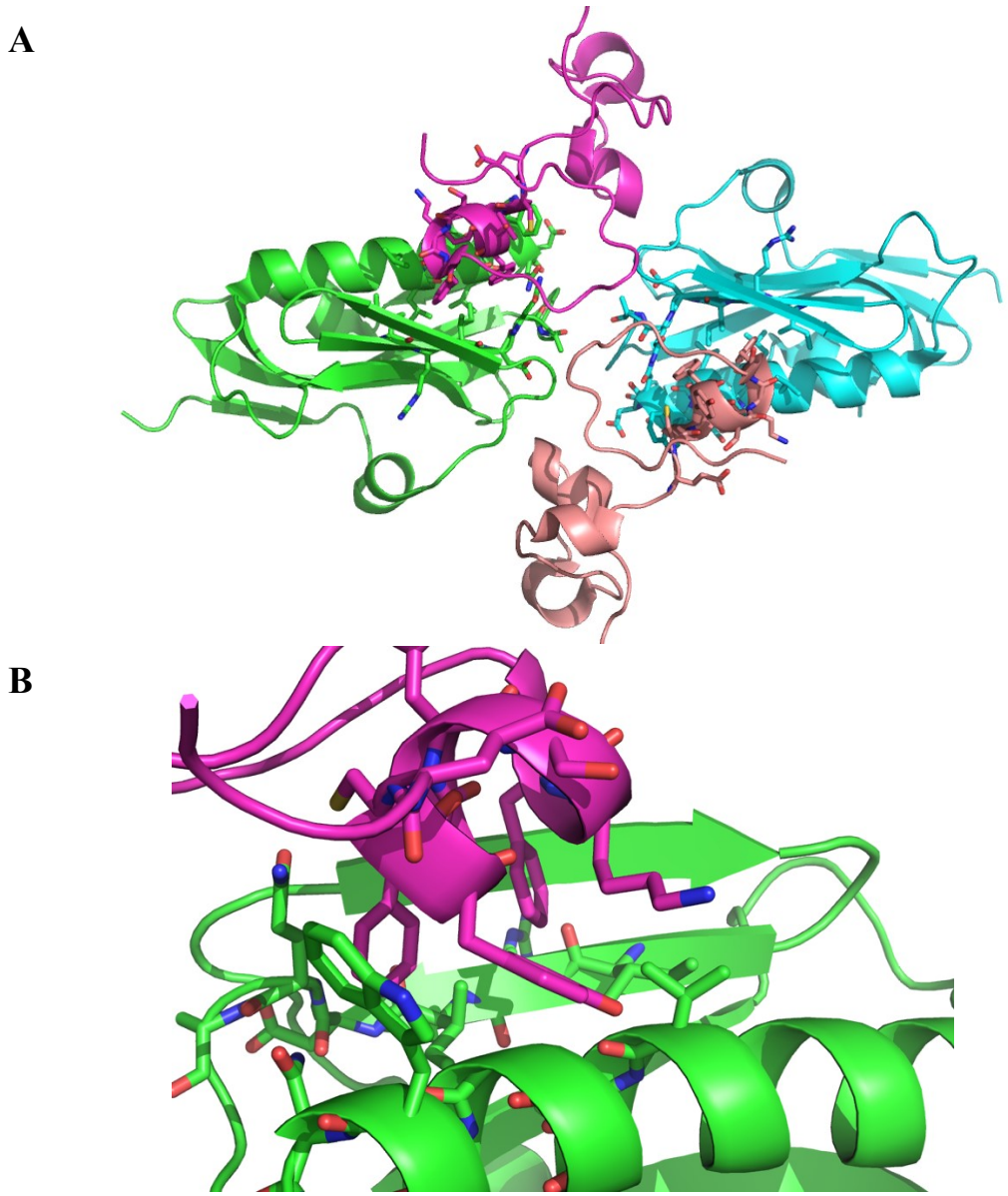


Figure 0-8 - Influenza A NS1 effector domain/CPSF30 complex

The NS1 Effector Domain forms a complex with CPSF30. In panel A, NS1 ED is colored in green and cyan; CPSF30 fragment is colored in magenta and peach. Panel B features a close look at the interactions between the CPSF30 fragment and the ED (PDB: 2RHK). The tryptophan observed at the binding site in Figure 0.4 is the same as the one observed here.

CHAPTER I

Diastereomeric control of helix capping

Chapter Abstract

D-amino acids provide functional advantages to engineered proteins and peptides including increased structural stability and half-life and the ability to design new structures and functions not available in the sequence space of traditional amino acids. Understanding the rules of secondary structure stability is necessary for optimal novel design of peptides and proteins. Prior research has demonstrated the increased stability provided by D-amino acids at the C-terminus of α-helices on a residue level. At a finer level, many forces come into play, including hydrogen bonding, helix dipole, and van der Waals' forces. By creating subtle D-amino acid substitutions in the C terminus of the Trp-cage helix we are able to explore the role of various intermolecular forces on the stability of the helix. Using the D-amino acid forms of serine, threonine, valine, and allo-threonine, we observed the effect of a sidechain-backbone hydrogen bond as well as the effect of helix terminus dehydration on thermostability of both the helix as well as the entire peptide.

INTRODUCTION

The fields of protein and peptide design are two young elements of the growing field of synthetic biology. With a heavy reliance on knowledge-based methodologies it can be difficult to design novel protein folds and structures that are not highly similar to natural proteins with documented structures.[5,6,72–74] Due to the availability of non-natural amino acids (NONAs), it is possible to design highly novel proteins and peptides; however, with a limited knowledge-base and without a strong understanding of the role of various interactions which define structure, the structure of a novel sequence is hard or impossible to predict.

Prior studies have demonstrated the benefit of D-amino acids on C-terminal caps by adding additional hydrogen bonds.[18–21] Techniques such as these have made it possible to design small peptides which normally have poor thermodynamic stability resulting in poor stability via protease degradation and weak activity.[7,36,75,76] Larger proteins often do not encounter such difficulty due to tertiary structure and hydrophobic core formation, adding significant thermostability. Current peptide design techniques used to avoid stability issues include covalent “staple” bonds forcing secondary structure formation, metal binding, or non-natural amino acids which include halides, however some of these approaches result in poor bioavailability, immune response, or other detrimental effects when applied to a therapeutic compound.[32,36]

Traditional knowledge of peptide stereochemistry derived from studies starting with Ramachandran in the middle of the 20th century are limited to knowledge based upon known natural proteins and peptides.[77] Further studies of di- and tri-peptides have increased the knowledgebase and recent structural data from peptides featuring non-

natural amino acids such as somatostatin have shown the positive effects of D-amino acids on activity.[3,19,20,78,79] Even still, the fine details of intra-residue atomic-level details have not been well studied where the coarse chemistry of the protein or peptide are identical with stereochemical differences in the sidechain. With threonine and isoleucine allowing additional stereochemical modifications, the true mirror image of the amino acid of the L-forms are the D-forms and D-allo forms exhibit a rearrangement at the second chiral center.[13]

We have set out to determine the roles of various intramolecular forces in a peptide/protein system by examining the impact of hydrogen bonds, dehydration, and dipole effects (Figure 1-2 - Effect of forces). Utilizing the D-amino acid forms of serine, threonine, valine, and allo-threonine, we are able to examine the impact of single atom changes on stability of an alpha-helix (Figure 1-3 - Substituted amino acids in 10 position). Due to the fundamentally similar nature of the sidechains of serine, threonine, valine, and allo-threonine, the minute change of each amino acid from one another is measurable. With the ability to analyze the differences between D-threonine and D-allo-threonine, we can measure the effects of a single methyl-hydroxyl switch in chirality and examine the subtle effects it has on protein secondary structure formation.

TC5b Trp-cage serves as a prototypical alpha-helix containing peptide.[27,80–82] Featuring a short sequence and a highly stable helix and secondary structure that is easily measurable using circular dichroism (CD), reliable quantification of the effects due to single amino-acid changes to its sequence is possible.[20,80,81,83,84] Its small size also offers the opportunity to examine atom-level interactions in a molecular dynamics simulation as well as by nuclear magnetic resonance. Furthermore, as the TC5b Trp-cage contains a

glycine at the helix terminus position 10 in a D-conformation, the ability to replace it with a D-amino acid should impart favorable energy gains by reduction of entropy. This will allow us to remove the energetic component of backbone entropy from the comparison.

MATERIALS AND METHODS

Sample preparation

Peptides were synthesized using solid phase 9-fluorenylmethoxycarbonyl (FMOC) chemistry at the Tufts University Core Facility (http://www.tucf.org). N- and C-termini were uncapped. Peptides were purified to >90% purity by reverse-phase high-performance liquid chromatography (HPLC) and products verified by mass spectrometry. Sequences in the study are as follows:

TC5b WT	NLYIQWLKD	G	GPSSGRPPPS
TC5b G10DAT	NLYIQWLKD	DAlloT	GPSSGRPPPS
TC5b G10T	NLYIQWLKD	DT	GPSSGRPPPS
TC5b G10S	NLYIQWLKD	DS	GPSSGRPPPS
TC5b G10V	NLYIQWLKD	DV	GPSSGRPPPS

Thermal and chemical denaturation

Peptides were dissolved in deionized water and the concentration of each were measured at 280 nm using an extinction coefficient of 6990 M^{-1} cm^{-1}. Peptide solutions at 50 µM concentration were prepared by diluting peptide stocks in 10 mM phosphate buffer, pH 7.0. CD experiments were performed on an Aviv model 420SF spectrometer (Aviv Biologicals, Lakewood, NJ). Ellipticity spectra were collected from 190 to 260 nm in 0.5 nm steps at 25 °C. A 1.0 nm bandwidth, 1 second averaging time, and 1 second

equilibration time were used. Molar ellipticity was calculated by multiplying raw values by the peptide concentration (50 uM), number of residues (20) and cell path length (1 mm).

For thermal denaturation, peptide solutions were diluted in 10 mM phosphate buffer (pH 7.0) to obtain a concentration of 50 uM. Ellipticity was measured at a wavelength of 222 nm with a bandwidth of 1.0 nm, temperature dead band of 0.15 °C, temperature equilibration time of 1 min, and averaging time of 2.0 s. Temperatures ranged from 0 to 80 °C in 2 °C steps. Temperature melts were collected for six concentrations of guanidine HCl (GnHCl) for each peptide: 0.0, 1.0, 2.0, 3.0, 4.0, and 5.0 M. Melting was demonstrated to be reversible after a period of cooling.

Temperature melt plots obtained for multiple GnHCl concentrations were fit to folded and unfolded baselines. Heat capacity was assumed to remain constant with changes in GnHCl concentration for each Trp-cage variant. The fraction folded, f(T), was calculated according to the method described by *Rodriguez-Granillo et al*.[20]

Nuclear Magnetic Resonance (NMR) spectroscopy and structure calculations

Spectra were collected on a Bruker Avance 800 MHz spectrometer at 25 °C. Samples were 1-2 mM protein in 10 mM potassium phosphate buffer, pH 7.0.

Experiments were conducted using unlabeled TC5b G10DT and G10DAT samples with 10% D_2O for frequency lock with 10 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) for internal chemical shift referencing. ^1H-^1H NOESY (nuclear Overhauser effect spectroscopy) spectra were collected using WATERGATE solvent

suppression and assigned using standard methods. ^{13}C-^{1}H HSQC (heteronuclear single quantum coherence spectroscopy) and ^{15}N-^{1}H HSQC spectra were collected using spectral widths of 11160 hz for ^{1}H, 14085 hz for ^{13}C and 2838 hz for ^{15}N for all HSQC experiments and a spectral width of 12820 hz for ^{1}H for all NOESY experiments. Mixing times of 50, 100, and 250 ms were used in the NOESY spectra.

Nuclear Overhauser effect intensities were obtained from NOESY spectra with constraints converted to distance and used in a simulated annealing protocol using the CYANA software package. Assignments and the restraint lists are included in the Supporting Information.

A total of 100 structures were created, out of which 10 structures were kept. For the TC5b G10DAT structure 400 restraints were used for structure calculation, which converged with zero violations. For the TC5b G10DT structure, 369 restraints were used for structure calculation which converged with only 6 van der Waals restraint violations which were likely due to ambiguous restraint assignments. The 10 lowest energy structures were deposited in the PDB.

Design of Trp-cage variants with ProtCAD

The computational design toolkit ProtCAD (Protein Computer Aided Design) was used to design and analyze D-amino acid containing Trp-cage variants.[85] The structural basis was the TC5b Trp-cage variant (PDB entry 1L2Y) where D-amino acids replaced the left-handed glycine at position 10. All rotamers of D-amino acids (threonine, allo-threonine, serine, and valine) were modeled in the 10 position and the lowest energy structure was used for simulation based upon an energy score computed using non-

bonded united-atom parameters from Amber 95 with dielectric electrostatic forces as described in Pike and Nanda 2015.[86] Hydrogen bonds were calculated using a sidechain Nε to backbone carbonyl oxygen distance cutoff of 3.5 Å.

Molecular Dynamics (MD) simulations

The atomic coordinates for the 1L2Y TC5b Trp-cage variant were obtained from the PDB. Structures generated in ProtCAD containing D-amino acids as well as control triplet peptides (GXG where X is a D-threonine, D-allo-threonine, D-valine, or D-serine) were prepared for simulation in the leap module of AmberTools 15, and simulations were performed for the variants using AMBER 14's PMEMD program accelerated with graphics accelerators.[28,87,88] Initial structures were immersed in a pre-equilibrated truncated box cell of TIP3P explicit water molecules, and counter ions were added to neutralize the systems. Protein atoms were described with the parm12SB force field parameterization. The protonation state of the titratable residues corresponded to the stable form at pH 7.0. Water molecules extended at least 10 Å from the surface of the proteins. Simulations were performed in the NPT ensemble (constant temperature of 300 K and constant pressure of 1 atm maintained using the Berendsen coupling scheme) employing periodic boundary conditions. The SHAKE algorithm was employed to keep bonds involving hydrogen atoms at their equilibrium length.

The systems were optimized and heated to 300 K and equilibrated for 5 ns. The structures were then simulated until the root-mean-square deviation (RMSD) as a function of time was stable for at least 100 ns (total simulation time was more than 200 ns for each). Converged simulation data was used for data analysis. RMSD, average

structures, hydrogen bond frequencies, and dihedral angles were calculated for each of the systems using VMD.[89] Cluster analysis was performed using the cluster tool in Chimera 1.9.[90]

RESULTS AND DISCUSSION

Design of the system

In order to isolate the effects of sidechain entropy, hydrogen bonding, and solvation, we chose to use the well-studied Trp-cage. Prior studies have shown that L-threonine is a suitable C-cap in left-handed helices. However, the mirror image and equivalent D-amino acid D-threonine was chosen along with D-allo-threonine in order to compare the effects of hydrogen bonding versus entropy and solvation effects.[21,91–93] D-threonine and D-allo-threonine consist of the same atoms with identical placement besides the configuration of a β-carbon bound methyl and hydroxyl group. D-serine contains the same β-carbon bound hydroxyl group, allowing for the same sidechain-helix backbone hydrogen bonding scheme but without the entropic effects of the adjacent methyl group (Figure 1-3 - Substituted amino acids in 10 position). D-valine was chosen to examine entropy reduction without the ability to form the sidechain-helix backbone hydrogen bond. As L-threonine and L-serine both demonstrate a high statistical preference for capping left-handed helices, we expected both to serve as a suitable cap.[21] We predicted D-allo-threonine would be even more beneficial due to the expected increase in steric clearance of its sidechain methyl group. Our model for this interaction scheme allows for a similar hydrogen bond as seen in D-glutamine in the same position.[20]

Capping diastereomer effects on stability

All D-amino acid containing variants show similar structure by circular dichroism. The same characteristic shape seen in all TC5b Trp-cage variants is present with a significant negative mean residue ellipticity (MRE) at 222 nm. D-amino acids that are able to hydrogen bond to the helix backbone carbonyls exhibit enhanced helicity at the characteristic 222 nm wavelength with an otherwise unaltered signal (Figure 1-4 - Substituting the G10 position of TC5b Trp-cage does not impart significant conformational change).

The melting temperature of the Trp-cage variants is increased greatly with hydrogen bonding variants (Figure 1-5 - Position 10 Sidechain to helix backbone hydrogen bonding significantly increases the thermal stability of Trp-cage). D-allo-threonine is the most thermally stable followed by D-threonine and D-serine. D-valine is the least thermally stable with a melting temperature even below that of wild-type Trp-cage.

Due to the stability of the three most stable variants, it is impossible to measure a baseline melted state even approaching 100°C. Therefore, it was necessary to partially unfold samples using guanidine HCl in order to determine the melted baseline. Using data from 0-5 M guanidine HCl, it is possible to fit the individual curves to generate an approximate melting curve of each variant.

Structure effects of substitutions

D-threonine and D-allo-threonine variants were examined by NMR to form an understanding of the structural basis for stability. Standard NH-HSQC and CH-HSQC

experiments were used to assign backbone and sidechain atoms. The backbone chemical shifts approximated the wild-type Trp-cage in other publications. Significant backbone chemical shifts were observed between D-threonine and D-allo-threonine variants near the substitutions, likely due to local environmental changes (Figure 1-6 - The backbone environment of G10DT and G10DAT is dissimilar near the 10 position residues).

NOESY derived distance measurements demonstrated a backbone structure similar to the wildtype Trp-cage structure with changes near the helix C-terminus. NOEs between the D-threonine and D-allo-threonine methyl group and nearby amino acids showed different relative distances suggesting considerably altered local orientation. Methyl carbon chemical shifts ruled out the predicted least-favorable rotamer.[94]

From these solution NMR structures, it is evident that both D-threonine and D-allo-threonine are very similar to the wild-type Trp-cage and the D-glutamine containing variant (Figure 1-8 – NMR ensemble of G10DAT and G10DT and Figure 1-9 - Lowest energy minimized structure derived from NMR data). In particular, the hydrogen-bonding scheme is similar between the two threonine variants with the primary distinguishing factor being the placement of the sidechain methyl group (Figure 1-7 - D-allo-threonine and D-threonine prefer rotamers which allow for sidechain-backbone hydrogen bond formation). In the case of D-allo-threonine, it is located as far away as possible from the helix. In D-threonine, it is partially occluding the helix causing partial dehydration. Still, the penalty of dehydration is small compared to the energetics gained from hydrogen bond formation and the reduction of entropic space the amino acid can explore.

Dynamic effects

D-threonine and D-allo-threonine variants greatly prefer one of three possible χ1 rotamers (Figure 1-10 - Molecular Dynamics simulation demonstrates considerable conformational sidechain χ1 preference allowing for sidechain-backbone hydrogen bonding and Figure 1-11 – D-allo-threonine exhibits a significant energy barrier to leaving a hydrogen-bound rotamer position). The dominant rotamer in each case is one that allows for hydrogen bond formation between the sidechain hydroxyl and the helix backbone carbonyls from either L7 or K8.

Unlike D-threonine and D-allo-threonine, D-valine and D-serine explore all three possible rotamer angles. D-serine does not feature a restrictive sidechain methyl group suggesting the increase in entropy slightly destabilizes the fold compared to threonine variants while the hydrogen bonding potential increases it. In the case of D-valine, the swap of a hydroxyl with another methyl group is energetically unfavorable and while sidechain entropy is reduced, the inability for a sidechain-backbone hydrogen bond and the forced desolvation of the helix terminus allow it to freely rotate in molecular dynamics simulation. Both of these results agree with melting temperature data and allow us to rank the subtle effects on stability.

D-valine provides sidechain entropy reduction but does not provide hydrogen bonding potential. It also is penalized with helix desolvation. D-serine provides hydrogen bonding potential and does not desolvate the helix; however, it does not provide a significant reduction of entropic space. D -threonine can hydrogen bond and provides entropy reduction but still partially desolvates the helix. D-allo-threonine can hydrogen

bond, the placement of the methyl group does not desolvate the helix, and still provides reduction of entropy.

All of these observations allow us to rank the effects in order of importance as follows: hydrogen bonding >> sidechain entropy > helix solvation.

CONCLUSIONS

D-allo-threonine is a functionally potent C-terminal helix cap. It allows for the formation of an additional hydrogen bond between the threonine sidechain to the i-2 or i-3 carbonyl while leveraging its ability to reduce the entropic space of sidechain motion. While D-threonine and D-serine are suitable C-terminus caps, they each have an elemental flaw: D-threonine by nature of structure must partially dehydrate the helix and D-serine does not reduce entropy as much as D-allo-threonine. In addition, it has been shown that the additional polarization of the threonine hydroxyl by the methyl group may impart a stronger hydrogen bond than serine can. It would be reasonable to assume that this slight energetic benefit helps to confer D-threonine its strength as a C-terminal cap.

With a growing arsenal of protein design techniques including D-amino acid helix caps, it is possible to design more stable proteins and peptides. D-glutamine and D-alanine are both excellent tools however there may be circumstances where the presence of a sidechain methyl group and a hydroxyl are more useful than a sidechain amide as exists in D-glutamine.

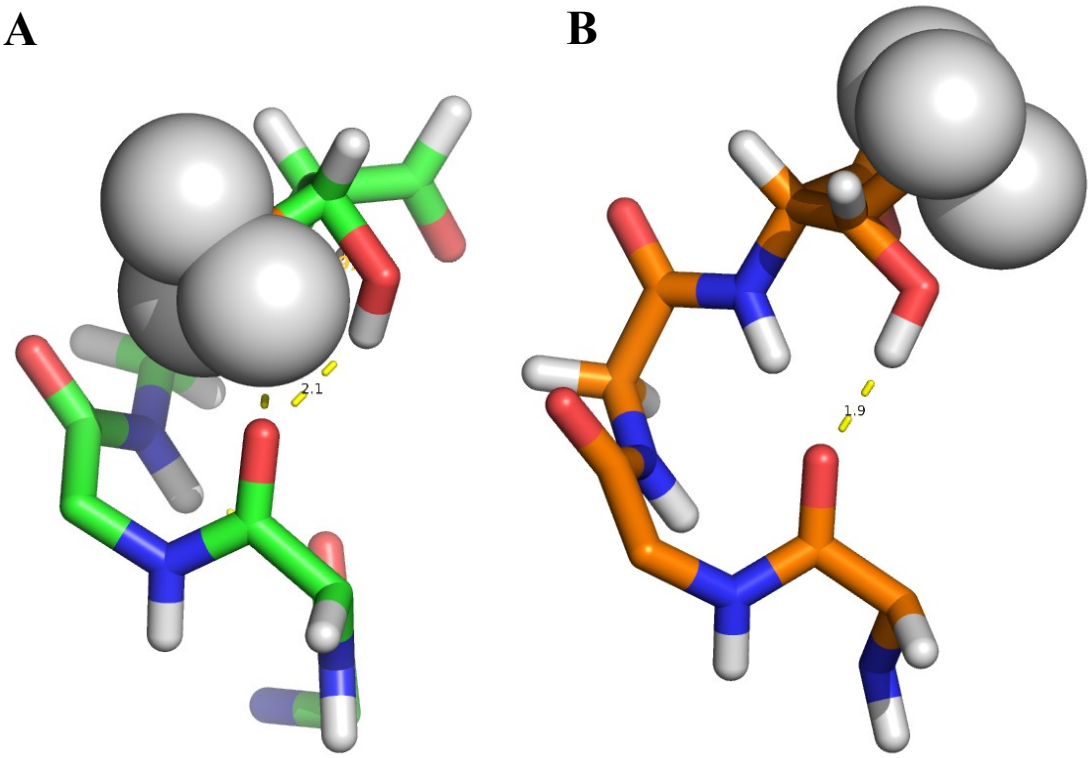


Figure 1-1 - Threonine C-caps in left handed helices

Prior work has demonstrated that L-Threonine is statistically favored as a cap in left-handed helices however we hypothesize that D-Threonine could add additional energetic stabilization by creating an additional hydrogen bond with the helix backbone. A. Illustration of PDB ID 1DS1 114-119 with D-Allo-Threonine methyl hydrogens shown as spheres. B. Model of 1DS1 with D-Threonine with methyl hydrogens shown as spheres.

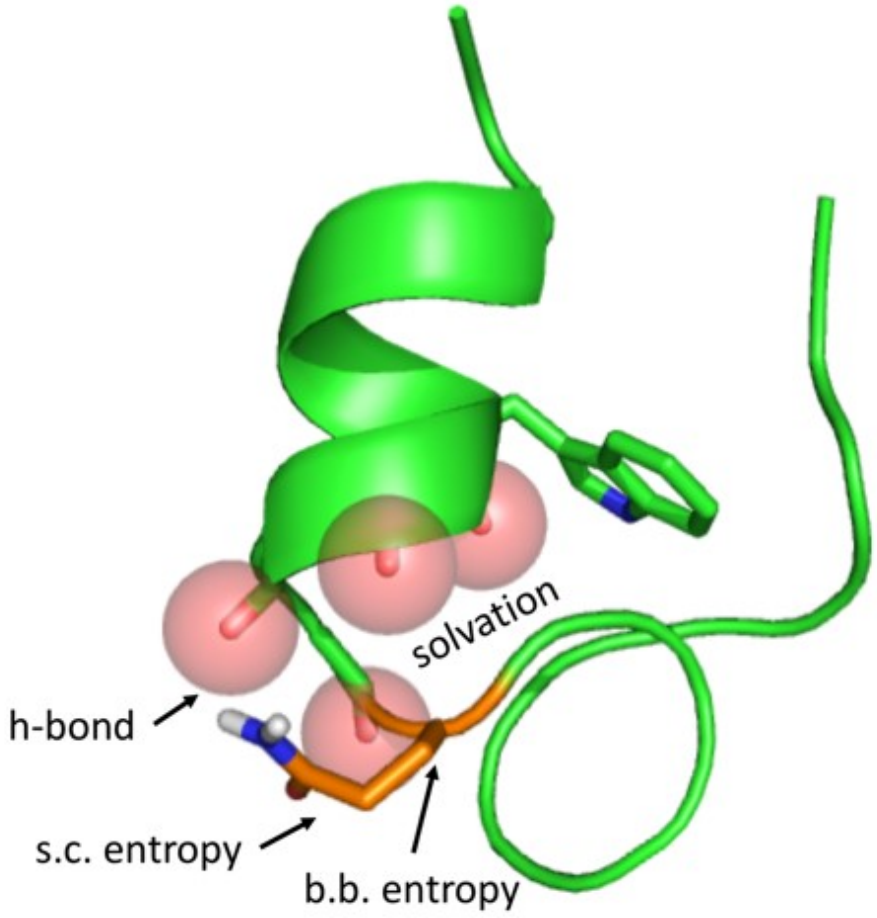


Figure 1-2 - Effect of forces

Various forces contribute to a stable helix cap. These forces include energy gained from hydrogen bonding, solvation, and entropy of the constituent amino acid sidechain and backbone.

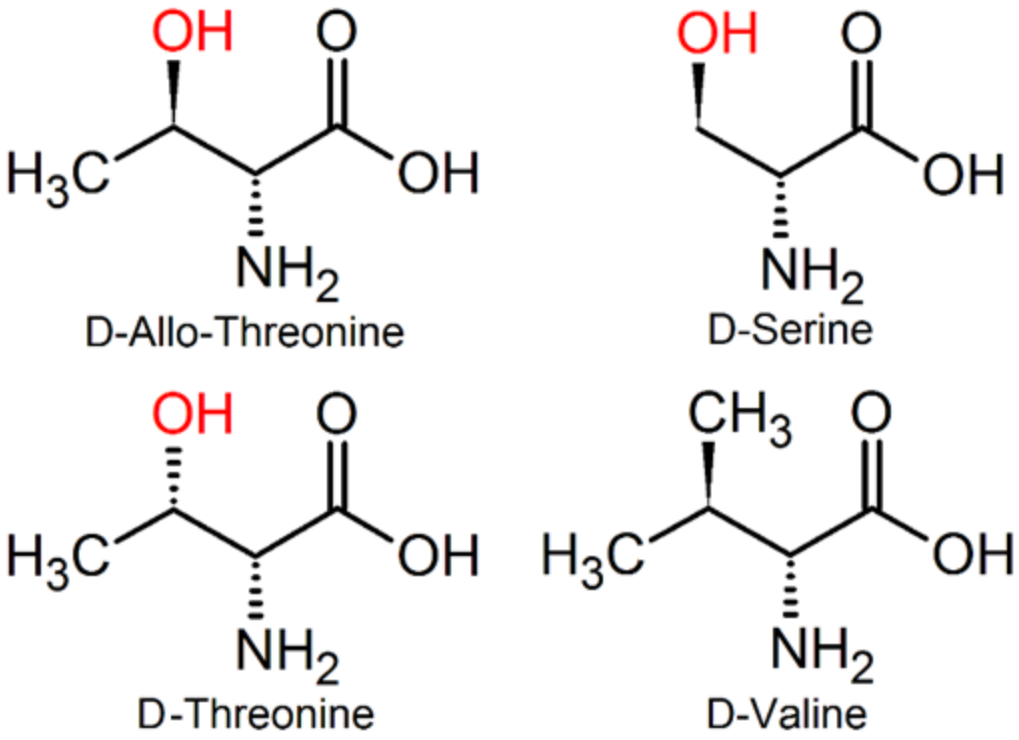


Figure 1-3 - Substituted amino acids in 10 position

The C-terminal Glycine 10 position is substituted with one of four D-amino acids to examine the impact of sidechain entropy, sidechain hydrogen bonding, and helix backbone hydration effects at the C-terminus of the helix in TC5b Trp-cage. All four amino acids contain the same β-carbon with additional groups allowing for steric entropic reduction, hydrogen bonding, or both.

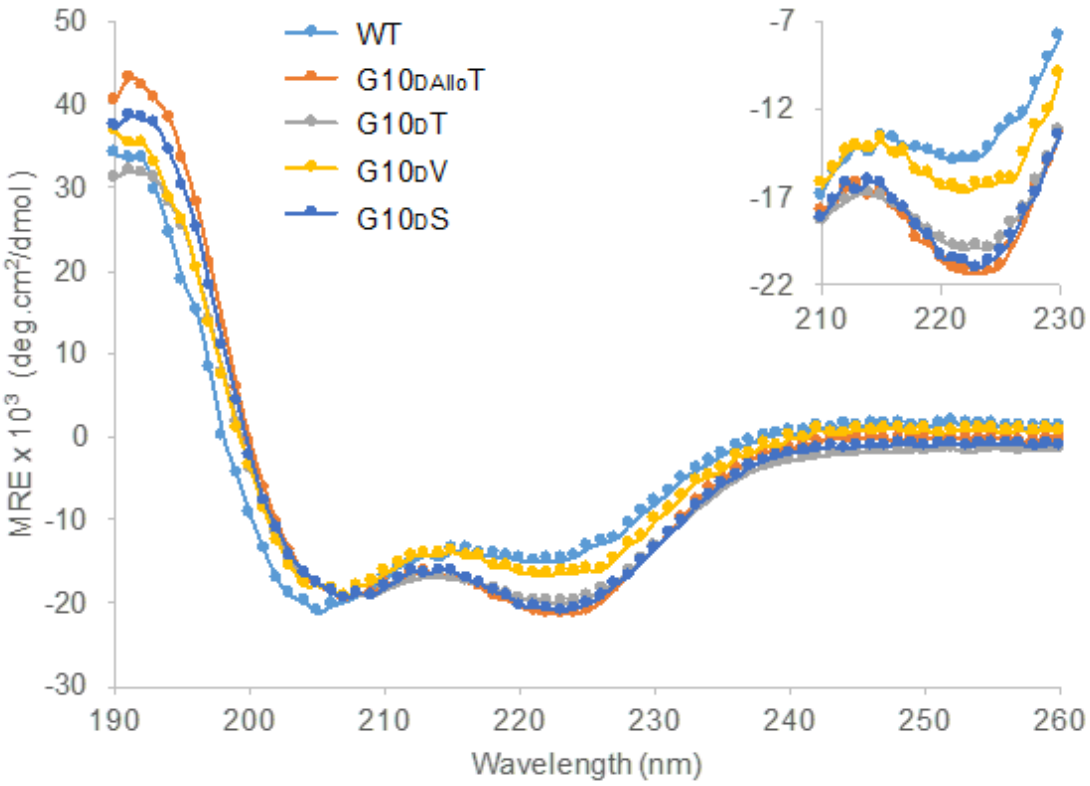


Figure 1-4 - Substituting the G10 position of TC5b Trp-cage does not impart significant conformational change

Variants featuring D-amino acids at position 10 still exhibit strong alpha helical character. D-Threonine, D-Allo-Threonine, and D-Serine all impart a stronger negative MRE at 222 nm suggesting a stronger helical character.

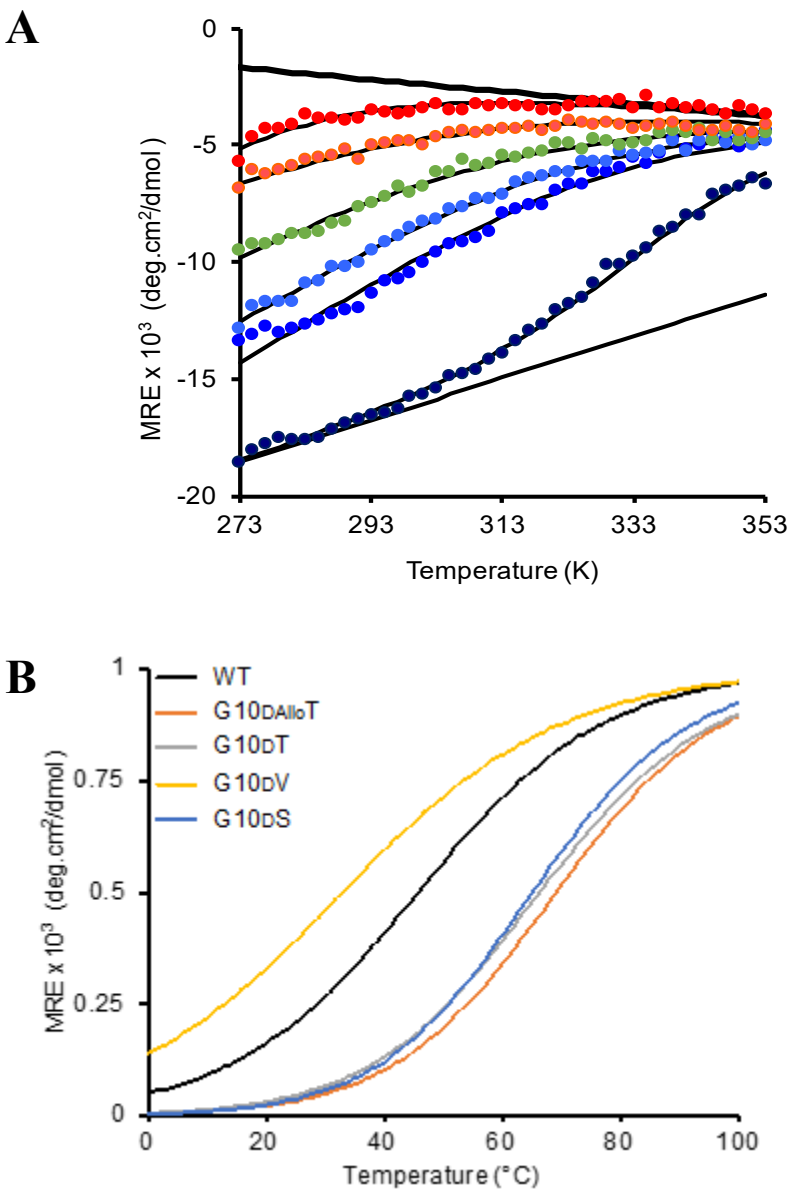


Figure 1-5 - Position 10 Sidechain to helix backbone hydrogen bonding significantly increases the thermal stability of Trp-cage

D-Allo-Threonine, followed by D -Threonine and D-Serine all increase the melting temperature of Trp-cage significantly above wildtype and D-Valine variants. A. Circular dichroism melt curves for D-Allo-Threonine in 0-5M GnHCl. B. Interpolated circular dichroism melt curves for all Trp-cage peptides.

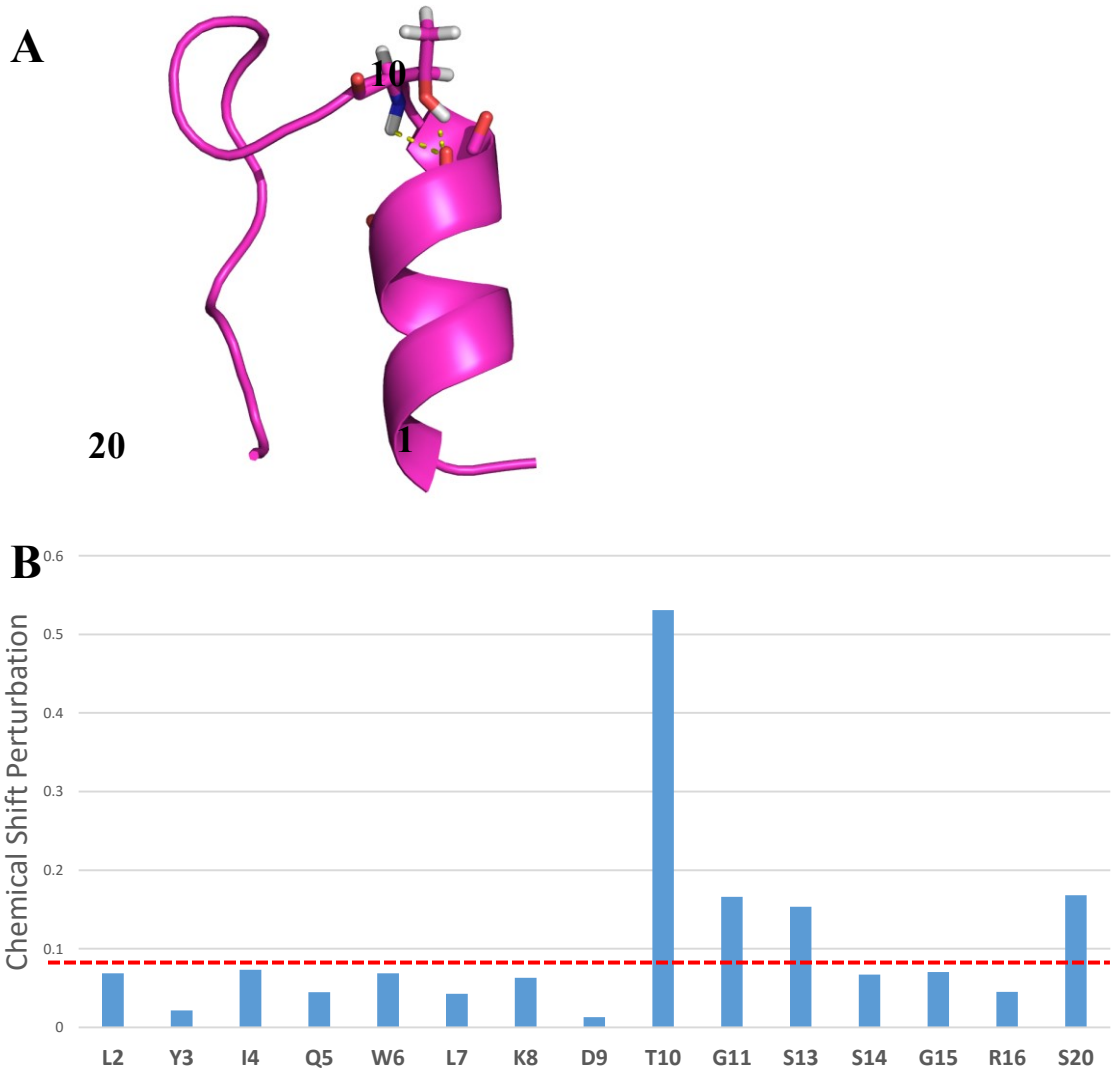


Figure 1-6 - The backbone environment of G10DT and G10DAT is dissimilar near the 10 position residues

Residues 10-13 all exhibit significant backbone amide HSQC chemical shifts. S20 also shifts significantly however it is loosely ordered. A. CYANA output of G10DAT with D-allo-threonine shown in sticks along with L7 and K8 carbonyls. B. Chemical shift perturbations of backbone amides of G10DT vs G10DAT.

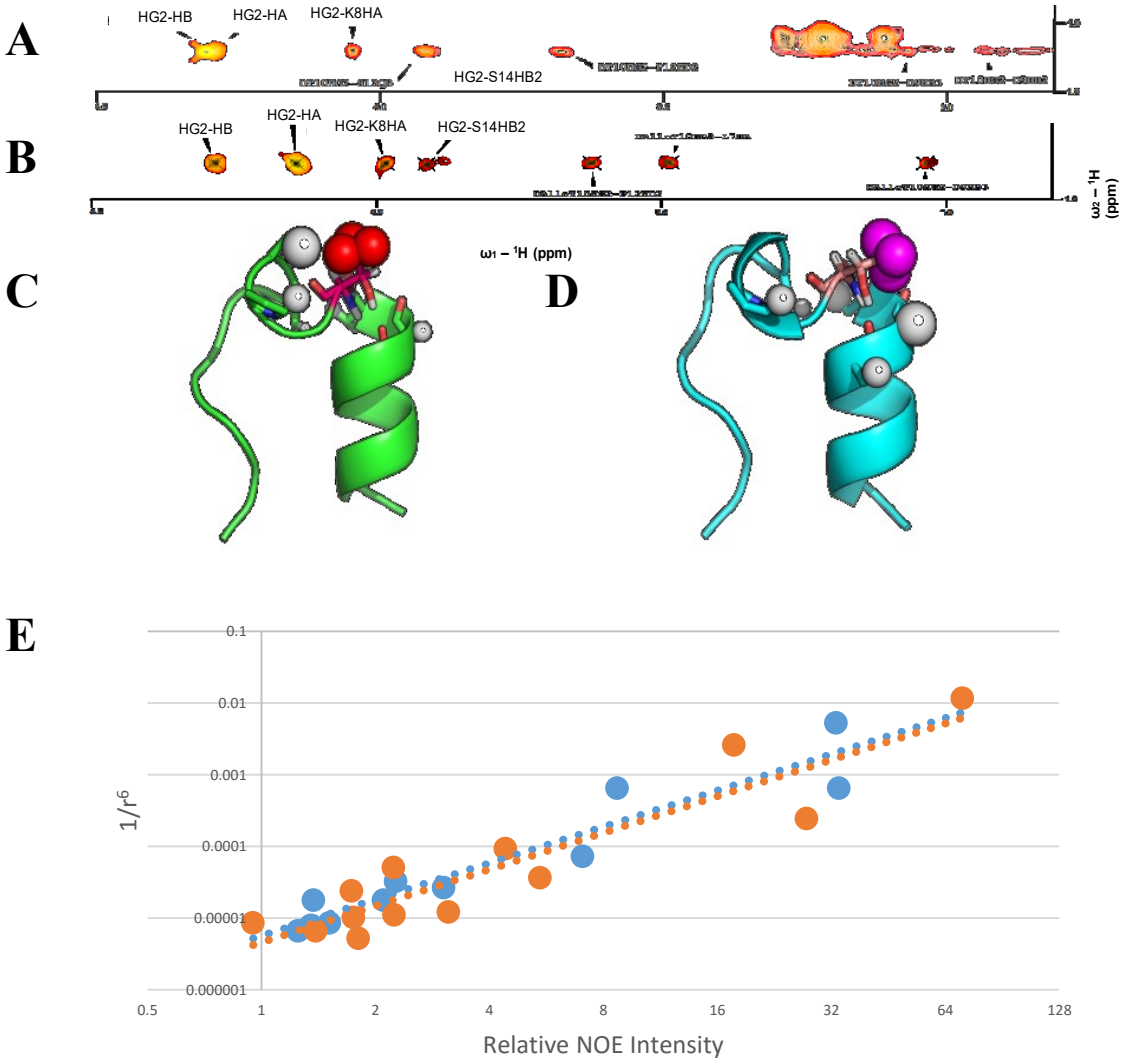


Figure 1-7 - D-allo-threonine and D-threonine prefer rotamers which allow for sidechain-backbone hydrogen bond formation

Analysis of the methyl NOESY spectrum demonstrates rotameric preference which allow for hydrogen bonding of the 10 position sidechain to the helix backbone carbonyls. In addition, the D-Allo-Threonine methyl prefers to be facing away from the helix allowing for the least helix carbonyl dehydration. A. G10DT methyl NOE crosspeaks. B. G10DAT methyl NOE crosspeaks. C. G10DAT CYANA generated lowest energy structure with methyl NOE crosspeaks shown as intensity-scale sized spheres. D. G10DT CYANA generated lowest energy structure with methyl NOE crosspeaks shown as intensity-scale sized spheres. E. Relative NOE intensity of G10DAT and G10DT vs $1/r^{6}$.

A B

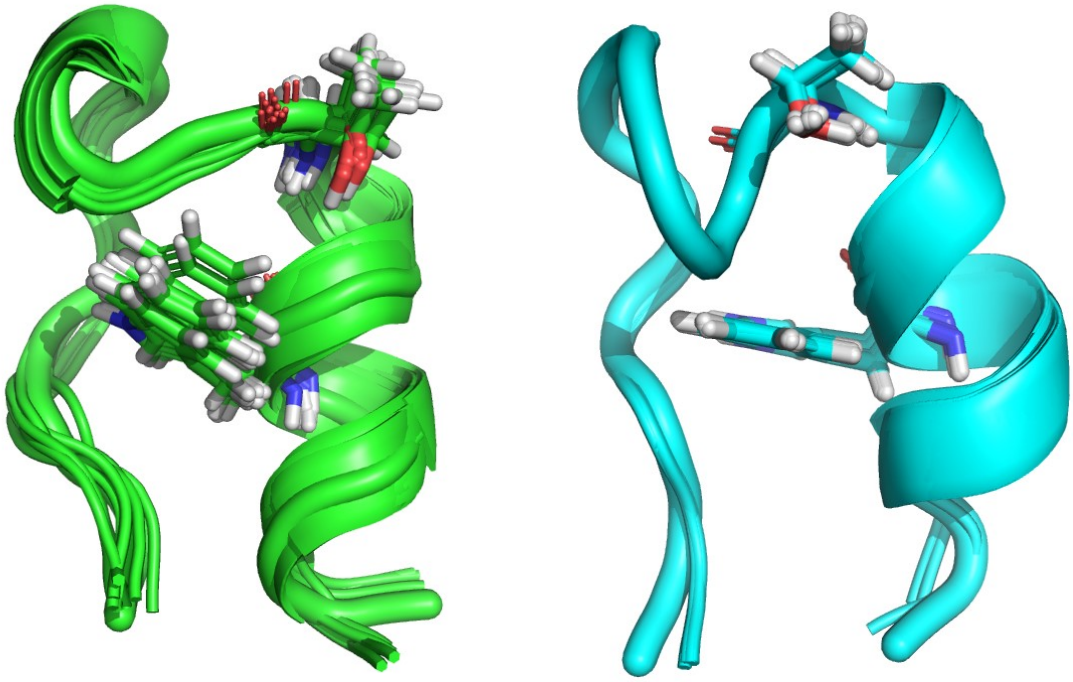


Figure 1-8 – NMR ensemble of G10DAT and G10DT

G10DAT (A, green) and G10DT (B, cyan) ensemble structures have identical Trp6 positioning indicative of a folded Trp-cage. In addition, the ensemble features identical 10 position rotamers.

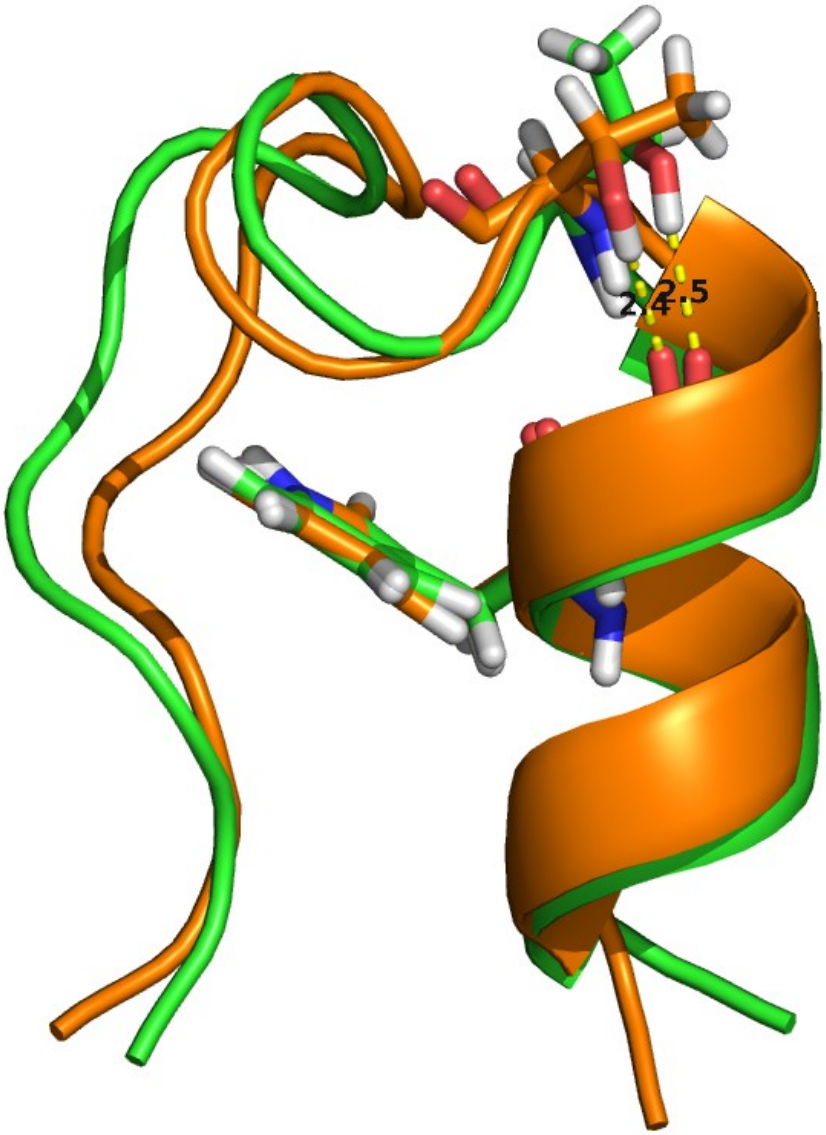


Figure 1-9 - Lowest energy minimized structure derived from NMR data

AMBER minimized structures of D-Allo-Threonine (green) and D-Threonine (orange) variants demonstrate the overall structure is similar with characteristic Trp6 placement. D-Allo-Threonine and D -Threonine sidechain hydroxyl groups are within hydrogen bonding distance (3.3 Å and 3.4 Å respectively) to Leu7 with only differences seen in the sidechain methyl.

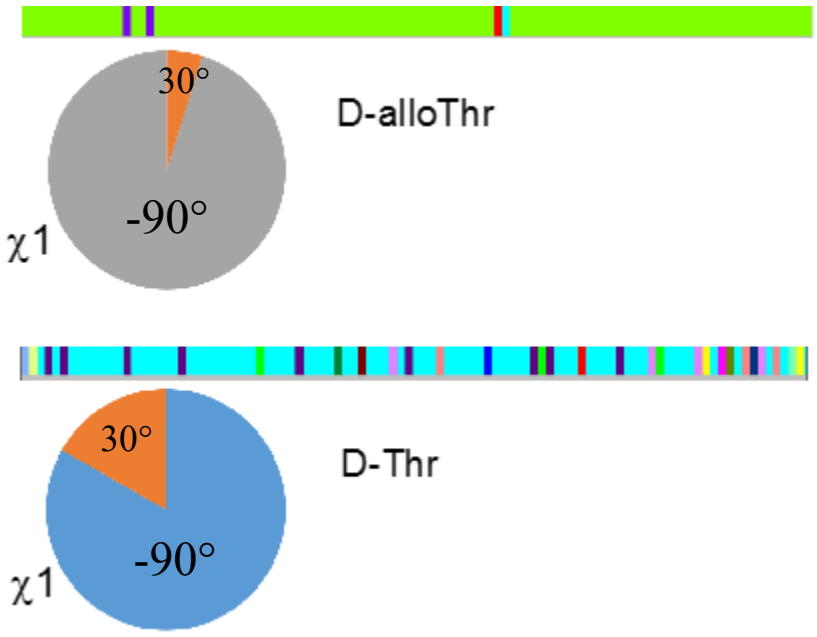


Figure 1-10 - Molecular Dynamics simulation demonstrates considerable conformational sidechain χ1 preference allowing for sidechain-backbone hydrogen bonding

Clusters of backbone conformations prefer one conformational state for a large majority of the simulation for both D-Threonine and D-Allo-Threonine with structural changes observed simultaneously along with χ1 occupying the less-occupied state. Horizontal bar colors indicate different structural clusters (as defined by all-atom RMSD).

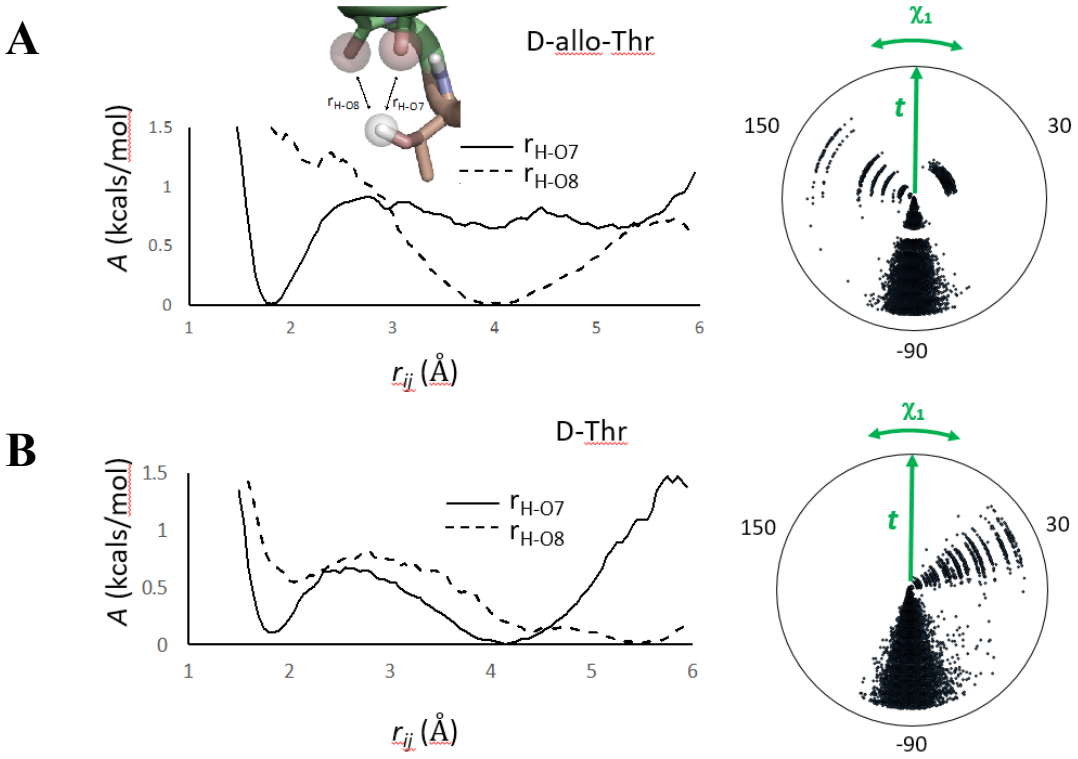


Figure 1-11 – D-allo-threonine exhibits a significant energy barrier to leaving a hydrogen-bound rotamer position

D-allo-threonine (A) has a larger energy barrier of approximately 0.8 kcal/mol to leave the -90° χ^1 rotamer where a hydrogen bond is shown to form. D-threonine (B) exhibits a significantly smaller energy barrier. In our molecular dynamics simulations, D-threonine can more freely rotate compared to D-allo-threonine.

	T_M (°C)	ΔG_U (25 °C) (kcal/mol)
WT	47 ± 3	0.9 ± 0.2
G10DAlloT	69 ± 0.5	1.8 ± 0.2
G10DT	66 ± 0.2	1.6 ± 0.1
G10DS	63 ± 0.1	1.7 ± 0.1
G10DV	33 ± 5	0.4 ± 0.3

Table 1.1 - Melting temperature and calculated change in free energy of folding

Melting temperatures were derived from three thermal denaturation CD scans at 222 nm each.

CHAPTER II

Influenza A non-structural protein 1 interaction partners and design of a short peptide inhibitor

Chapter Abstract

The non-structural protein 1 of the influenza A virus (NS1A) provides a multi-faceted approach for inhibition of the host immune response. The C-terminal effector domain (ED) binds several partners, including the host factor CPSF30, and is responsible for homodimerization with a second NS1A molecule. NS1A is highly conserved across influenza subtypes with the interface for these C-terminal ED interactions sharing significant identity. Tryptophan 187 has been previously demonstrated to be important to both binding events. Due to its strong conservation and role in important interactions which inhibit the host anti-viral response, we have examined the interface using fluorine ^{19}F NMR and designed a short peptide inhibitor to inhibit both its homodimeric and heterodimeric interactions.

INTRODUCTION

Many elements of the influenza A virus should stimulate a strong host immune response if not for clever immune inhibition and disguise tricks. The intermediate formation of a foreign double-stranded RNA (dsRNA) intermediate during viral replication usually would be sufficient for immune-triggered apoptosis and a rapid end to the viral life-cycle. However, due to several interdependent-yet-harmonious functions of the viral non-structural protein 1 (NS1), the virus can evade the natural innate immune system functions of the body.[55,95,96] Because of the multi-functional nature of the NS1 protein, it is often considered a viral "Swiss army knife." Analysis of the coding sequence for the NS1 protein has very high sequence conservation, likely due to a segment of shared coding region with the non-structural protein 2 (NS2) which has a different reading frame.[56,60,97–99]

The NS1 protein is composed of two domains: one located at the N-terminus and shown to bind double-stranded RNA nonspecifically (RNA-binding domain, or RBD) and one located at the C-terminus and shown to have several protein interaction partners (effector domain, or ED).[55,98] The protein has been shown to form higher-order polymeric structures involving interactions between at least two interfaces.[59,65,100] The RBD dimerizes and forms a strong dsRNA-binding dimer and the ED has been show to homodimerize in various ways. In addition to binding itself, the ED binds a 30-kDa host protein termed cleavage and polyadenylation specificity factor (CPSF30), a subunit of the multi-protein CPSF complex necessary for maturation of host pre-mRNAs including immune-related proteins [e.g. interferon (IFN)]. [60,61] Both CPSF30 and the ED

homodimer involve the same binding surface with a key tryptophan residue (W187) necessary in the formation of both complexes (Figure 2-1 – NS1A binds CPSF30 helix near tryptophan 187 and Figure 2-2 – NS1A binds itself as a homodimer along the same interface as CPSF30). This binding site for both homodimer and CPSF30 binding has been previously proposed as an antiviral drug target.[101]

Models in previous studies hypothesized a complex in which a dsRNA, presumably an intermediate produced during viral replication, is protected from the cell immune detection sensors by a sheath of NS1 proteins connected together by the ED dimer interface.[102] Cooperative NS1-RNA-binding interactions may also be important in other NS1 functions involving ribonucleoprotein (RNP) complexes (Montelione, G.T., personal communication). Studies of engineered transgenic influenza viruses with single point mutations in these NS1 protein interfaces have significantly reduced virulence.[60,96,102]

By interrupting the ability for the NS1 protein to form complexes with itself and with other binding partners, it may be possible to disrupt the immune-suppressive activities of the virus. In this study, we design short peptides to mimic the CPSF30 protein partner with the goal of disrupting the NS1 ED homodimer and NS1-CPSF30 heterodimer.

In our study, we examine the dynamics of the ED in the presence of F2F3 using different ^{19}F isotope labeled amino acids allowing us to use two different probes to independently isolate the effects of binding from the point of view of each molecule. In addition, we studied the effect of our design peptide on the ED homodimer system as well as the ED-F2F3 complex using the same probes.

MATERIALS AND METHODS

Protein sample preparation

Constructs of wild-type influenza A/Udorn/307/1972 NS1(85-215), referred to as ED or effector domain, were cloned with a C-terminal hexa-histidine (6xHis) affinity purification tag in a pET21_NESG vector.[103] The construct consisting of the [S94]-F2F3 fragment of CPSF30(60-120), referred to as F2F3, was fused with a 6xHis tagged SUMO protein in the pSUMO vector (LifeSensors) modified to be resistant to ampicillin. The constructs were expressed, and purified following the standard operating protocols of the Northeast Structural Genomics Consortium (NESG). The C-terminal 22 amino acids of the full-length NS1 protein are unstructured and lead to insolubility and were thus not included in our constructs. A single mutant protein (K110 to A) referred to as K110A ED was made using the QuikChange site-directed mutagenesis kit (Stratagene) along with previously developed primers and was verified by sequencing.[59] The expression was performed in the *Escherichia coli* strain BL21(DE3)-Gold (Agilent Technologies) containing the rare tRNA codon-enhanced pMgK plasmid and isopropyl β-D-thiogalactoside (IPTG)-inducible T7 polymerase. Cultures were grown in 2-liter baffled flasks at 37°C in MJ9 minimal medium supplemented with ampicillin and kanamycin.[104] For the production of ^{13}C- and ^{15}N- double-labeled samples, the MJ9 medium contained D-[U-^{13}C]-glucose and $(^{15}NH_4)_2SO_4$ as the sole carbon and nitrogen sources. For 5-^{19}F-Tryptophan (5FW) incorporation, cultures were grown until an A_{600} of ~0.5 units, cooled on ice with addition of 50 mg/l 5-fluoro-DL-tryptophan (Sigma), and the incubated at 17°C for 1 hour before induction with 1 mM IPTG. Protein expression was carried out

overnight at 17°C with shaking at 225 rpm. Cultures were subsequently centrifuged and decanted with pellets stored at -80°C until purification.

All expressed constructs were purified by immobilized nickel ion chromatography followed by size exclusion chromatography in phosphate-buffered saline (PBS) [137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, and 1.8 mM monopotassium phosphate with pH adjusted using HCl or NaOH to 7.4] or low salt pH 8.0 buffer [50 mM Tris-HCl (pH 8.0), 30 mM NaCl, 2.5 (v/v) glycerol, 10 mM dithiothreitol (DTT)]. Pellets were resuspended in 30 ml of nickel affinity column binding buffer [50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 40 mM imidazole, 1 mM Tris(2-carboxyethyl)phosphine, and 0.02% NaN_3] followed by sonication, and clarification by centrifugation at 28,000 x *g* for 60 minutes. Proteins were then loaded on a gravity flow column containing 4 ml of Ni-NTA Superflow resin (Qiagen), washed using 20 mL of the binding buffer, and eluted in 10 ml of nickel affinity elution buffer [50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 300 mM imidazole, 1 mM Tris(2-carboxyethyl)phosphine, and 0.02% NaN_3]. Elution fractions were then further purified using an AKTAexpress (GE Healthcare) system with a HiLoad 26/600 Superdex 200 gel filtration column equilibrated with the final buffer. For the SUMO-fusion F2F3, cleavage of the 6xHis and SUMO tag was performed using an in-house expressed SUMO protease Ulp1 aliquot and sample incubated overnight at 4°C with cleavage verification by SDS-PAGE.[105] Samples were concentrated by ultrafiltration (Millipore).

Peptide design

The peptide was designed using D-amino acid stabilizing caps derived from previous studies.[20,21] The majority of the sequence was derived from the wild-type CPSF30 sequence in order to bind the same site. The N-terminal capping motif used was a position 1 D-asparagine, which has been shown to stabilize α helices by forming stabilizing hydrogen bonds.[21] The C-terminal capping motif used was a final position glycine proceeded by a D-glutamine, also shown to stabilize α helices by forming stabilizing hydrogen bonds.[20] The final peptide sequence is D-*N*YFYSKF D-*Q*G where *N* and *Q* are the D-form amino acid and is referred to as Pep2K.

The peptides were modeled using the protein design software derived from *ProtCAD* and simulated in molecular dynamics for 200 nanoseconds to examine stability.[85] Peptides were synthesized using solid phase 9-fluorenylmethoxycarbonyl (FMOC) chemistry at the Tufts University Core Facility (http://www.tucf.org). N and C termini were uncapped. Peptides were purified to >90% purity by reverse-phase high-performance liquid chromatography (HPLC) and products verified by mass spectrometry.

^{19}F NMR of the complex

1D ^{19}F NMR was performed locked and at 25°C on a Varian INOVA 500 MHz spectrometer equipped with a room temperature 5 mm ^{1}H/^{19}F switchable probe at a frequency of 470.18 MHz or a Bruker AVANCE III 600 MHz equipped with a cryogenic 5 mm ^{1}H/^{13}C/^{15}N/^{19}F switchable probe at a frequency of 564.09 MHz. All ^{19}F NMR

spectra were acquired using VNMRJ 2.1B or TopSpin 3.2 and referenced to external neat trichlorofluoromethane. Typical 1D ^{19}F NMR acquisition parameters were as follows: 20,000 Hz sweep width (42.5 ppm), a 0.35 s acquisition time, a 1 or 5 s relaxation delay time, and a 5.0 μs 90° pulse length. 1D ^{19}F NMR spectra were processed with 75 Hz exponential line broadening and displayed using ACD NMR.

RESULTS AND DISCUSSION

CPSF30 disrupts the ED homodimer

The ability for the influenza A NS1 to bind the host CPSF30 protein has been demonstrated in the past with observed multimerization in solution NMR and crystal conditions.[60] The effector domain alone has been previously shown to form a dimer in low salt conditions (30 mM NaCl) with a dimer K_D of between 12 and 35 μM with significantly lower dimer species observed in physiological concentrations.[59] We also studied heterodimer complex formation using the F2F3 fragment of CPSF30, corresponding to the second and third zinc fingers of the CPSF30 protein. Using ^{19}F NMR, we have demonstrated that under physiological salt conditions, the dimerization constant of ED and the F2F3 of CPSF30 fragment is on the order of 60 μM K_D, an interaction significantly stronger than the ED homodimer (Figure 2-3 – CPSF30 disrupts the NS1 ED homodimer).

Two different experiments were used to evaluate F2F3 binding to the ED. The second and third zinc fingers of the CPSF30 protein were used (F2F3) in titration to the ED. In one experiment in low salt concentrations, unlabeled F2F3 was titrated with 5-fluoro-tryptophan (5FW) ED. In this experiment, the characteristic and previously reported broad W187 peak shifts and narrows with increasing concentration of F2F3 indicating a transition from ED homodimer to ED-F2F3 heterodimer. Hence, the disruption of homodimerization is observed from the context of the ED protein. Splitting of the W102 and W203 peaks occurs due to the non-helix elements of the F2F3 zinc

finger weakly binding to the ED. In high-salt concentration conditions, this splitting is abrogated.

In our experiments in various salt concentrations, it appears that in low-salt concentrations, the W187 peak is broad and poorly defined. As salt concentrations increase, individual small peaks appear and at high-salt concentrations, the W187 peak moves from several small peaks to a single, highly intense and narrow peak. Our analysis suggests that at low-salt concentrations, the flexibility of the W187 and the relative orientation of the dimer is not fixed into distinct states. As salt concentrations increase, transiently stable conformations are reduced and instead the effector domain homodimer adopts fewer, more stable conformations which resolve distinct peaks. Finally, with high-salt concentrations, the relative ratio of dimer to monomer is significantly lower with most of the signal from the W187 coming from monomer ED.

In another experiment, we titrated 4-fluoro-phenylalanine (4FF) labeled CPSF30-F2F3 fragment into the 5-fluoro-tryptophan (5FW) labeled ED (Figure 2-4 – CPSF30-F2F3 and NS1A ED exhibit chemical shift perturbations upon binding). In this experiment, the chemical shift changes of each component can be independently evaluated. The new chemical shift peaks contributed by the 4FF labeled F2F3 fragments do not overlap with the 5FW labeled ED allowing for clear context-independent consideration. With the addition of 4FF F2F3, the characteristic 5FW peak corresponding to W187 narrows. In addition, with increasing concentration of 4FF F2F3, new peaks are evolved downfield with two new significant peaks. We evaluated 4FF F2F3 alone without any ED in order to identify chemical shift changes.

Designed peptide disrupts the ED homodimer

Similar to our experiments titrating F2F3 into a solution containing 5FW labeled effector domain, we titrated our designed peptide, Pep2K, and observed chemical shift changes in low salt concentrations (Figure 2-5 – Designed peptide binds near tryptophan 187 of the NS1 ED). It is clear that the peptide disrupts ED homodimer population and the characteristic broad dimer peak moves to an extremely intense, narrow monomer peak in the presence of peptide. We independently verified the binding of the peptide to ^{13}C, ^{15}N labeled ED without the 5-fluoro-tryptophan incorporation in order to validate binding in the context of non-fluorinated protein (Figure 2-6 – Peptide induces a chemical shift of the W187 backbone). A strong slow-exchange phenomenon was observed with W187 peak splitting.

The W187 peak intensifies considerably in the presence of peptide suggesting a change from a series of homodimer states to one of a new, single heterodimer state.

Designed peptide disrupts the ED-CPSF30 complex

After titrating in 4FF F2F3 into the 5FW ED solution, we titrated Pep2K and evaluated chemical shifts during the titration (Figure 2-7 – Designed peptide inhibits CPSF30-F2F3 binding to NS1A ED). We observed a significant change in intensity of the W187 peak analogous to that observed when we titrated the peptide without F2F3. In addition to the chemical shifts in the effector domain, the broad peak from 4FF F2F3 decreased in intensity until it completely disappeared, concomitant with a new partially-phased peak increasing in intensity. ***Hence the designed peptide can successfully***

compete with F2F3 for NS1 ED binding, and provide competitive inhibition of this biologically important interaction.

CONCLUSIONS

The NS1 protein exhibits a multifunctional role with a broad array of partners. Two critical interactions that have been reported are the homodimer interaction which may contribute to cooperative binding and in its interactions with double-stranded RNA. The same interface is involved in heterodimer formation between the ED and CPSF30, which functions to disrupt of interferon pre-RNA processing.[60] Both dimer forms utilize essentially the same surface of the NS1 ED, which is therefore a versatile target on NS1 for an inhibitor compound.

F2F3 is a tandem zinc finger fragment of CSPS30. We have demonstrated that the ED-F2F2 interaction is stronger than the ED homodimer and that both can be inhibited by blocking the interface adjacent to W187. Our designed peptide disrupts both interfaces by competitively inhibiting both the homodimerization of the ED, and heterodimer formation with F2F3.

Our design philosophy based upon prior studies involving D-amino acids to stabilize α-helices has served as an effective diagnostic tool to examine the binding interface. It is possible that through further engineering, peptide inhibitors such as this may be able to inhibit the influenza anti-immune interactions necessary for viral replication.

Still, many limitations of peptide drugs remain. Various unanswered questions regarding peptide delivery present roadblocks impeding the movement from the laboratory into the clinic. Difficulties surrounding delivery, stability, and immunogenicity, have provided researchers with a wide range of study. In the case of systemic diseases, the larger size of peptides compared to small molecules prohibits the usual methods of cellular entry. As standard L-amino acid peptides are often less structurally stable than proteins, they are susceptible to proteolytic degradation and misfolding. Due to the protein-like features, peptides also may trigger immune detection and targeting.

In the case of the influenza A NS1 binding peptide described in this study, issues with delivery are partially resolved by administration directly to the airways. Since influenza infects epithelial tissues in the lungs and air passageways, an inhalant-based delivery approach may be used to specifically target the proper affected tissues. In addition, the hydrophobic and positively charged nature of the peptide imparts a theoretical benefit for endocytosis-mediated import into cells. With the presence of two D-amino acids, the majority of proteases are unable to access bonds to induce cleavage presenting increased stability. Finally, since the peptide is designed to have increased structural stability along with a short nine amino acid long sequence, the ability to be recognized reliably by the immune system is reduced. While we have no studied the factors listed regarding delivery, stability, and immunogenicity, we believe D-amino acids serving as structural caps may help reduce these factors impact on activity.

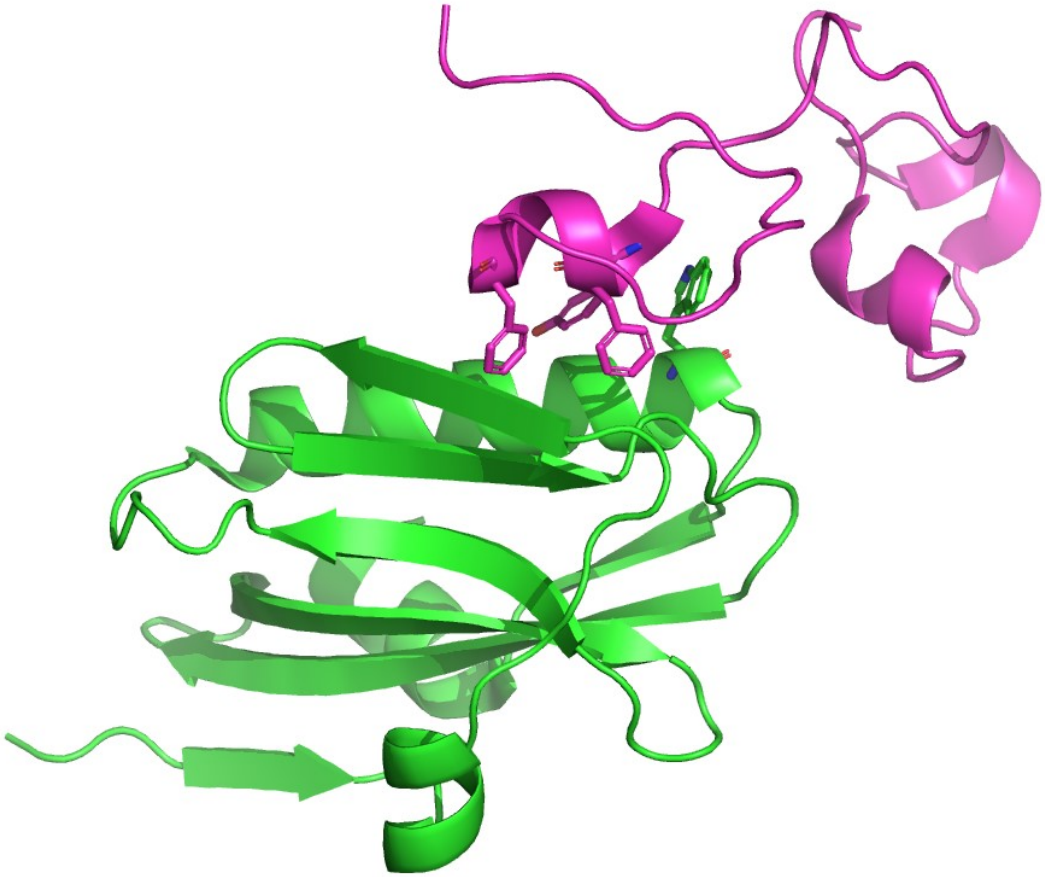


Figure 2-1 – NS1A binds CPSF30 helix near tryptophan 187

The influenza A NS1 effector domain (green) binds a short helix from the F3 zinc finger of CPSF30 (magenta). Critical interactions involve residues (shown in sticks) Y97 and F98 as well as F102 of the CPSF30 along the cleft created by the effector domain helix-sheet interface which features W187 (PDB 2RHK).

Figure 2-2 – NS1A binds itself as a homodimer along the same interface as CPSF30

The influenza A effector domain (green and cyan) forms a homodimer at the same helix-sheet cleft as seen in figure 2-1. A key tryptophan amino acid at position 187 contributes to form the homodimer interface (PDB 3EE9).

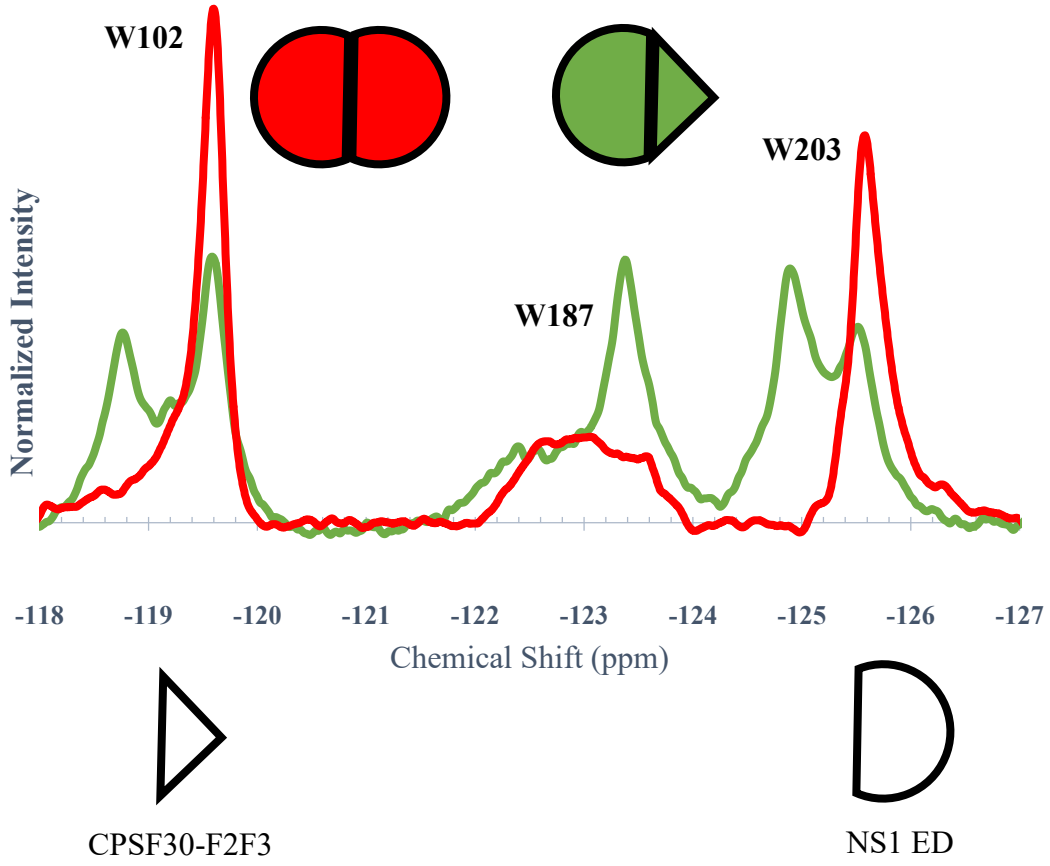


Figure 2-3 – CPSF30 disrupts the NS1 ED homodimer

Adding F2F3 zinc fingers of CPSF30 induces chemical shift changes in the W102, W187, and W203 peaks of 5-fluoro-tryptophan labeled NS1A ED. Splitting of the W102 and W203 peaks is induced upon binding while W187 moves from a broad and dynamic peak towards a narrow peak in low-salt buffer (30 mM NaCl). Red – 150 µM 5FW-labeled NS1A ED. Green – 125 µM 5FW-labeled NS1A ED in the presence of 125 µM unlabeled CPSF30-F2F3.

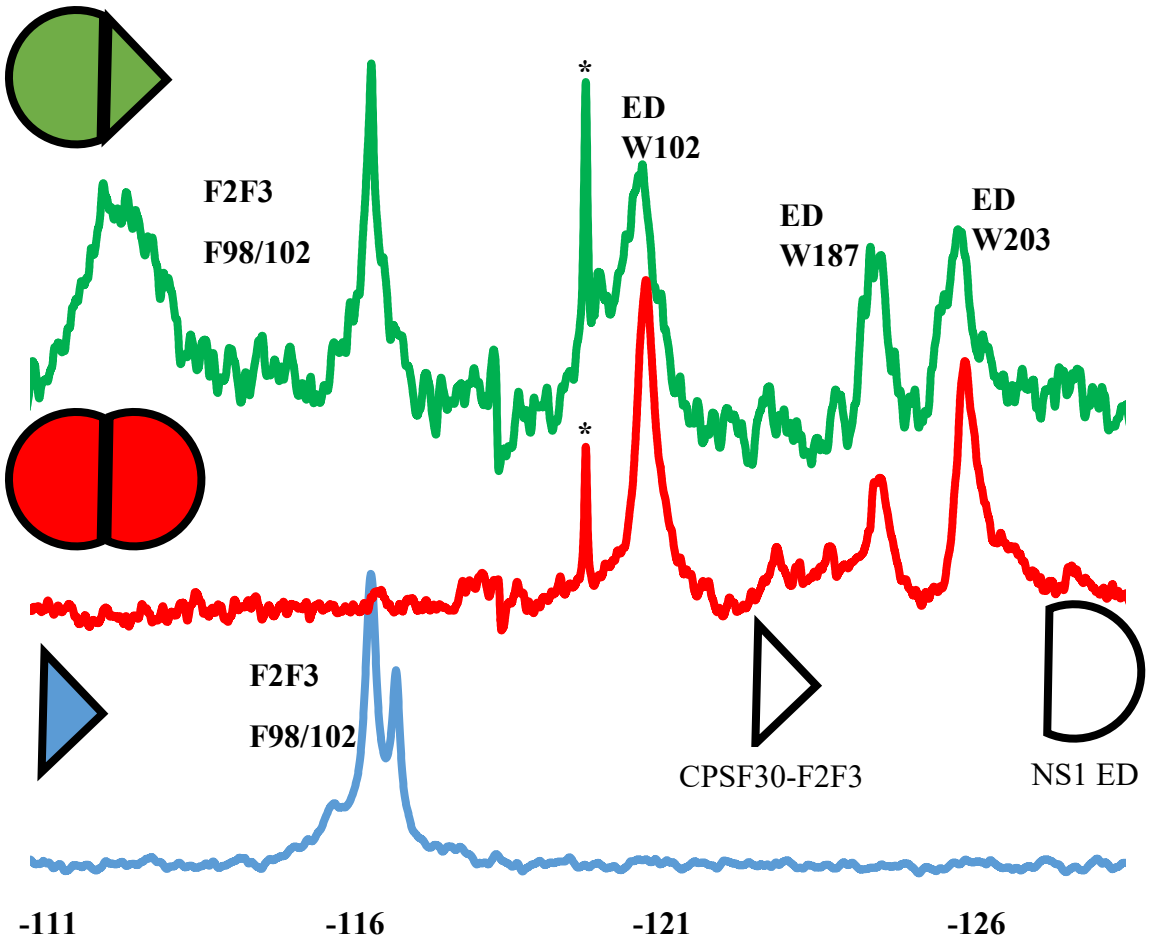


Figure 2-4 – CPSF30-F2F3 and NS1A ED exhibit chemical shift perturbations upon binding

A mixture of 1:1 ratio 5-fluoro-tryptophan labeled NS1A ED with 4-fluoro-phenylalanine CPSF30 zinc fingers F2F3 induces chemical shift perturbations to both the ED W187 as well as F2F3 in phosphate buffered saline. The use of two independent probes, 4-fluoro-phenylalanine on F2F3 and 5-fluoro-tryptophan on the NS1A ED allows independent observation of binding. Upon binding to F2F3, W187 narrows and intensifies. Likewise, One F2F3 peak shifts considerably into a new, broad peak downfield. Blue – 100 μM CPSF30-F2F3, Red – 100 μM NS1A ED, Green – 100 μM CPSF30-F2F3 and 100 μM NS1A ED. * - protease inhibitor purified with NS1A ED.

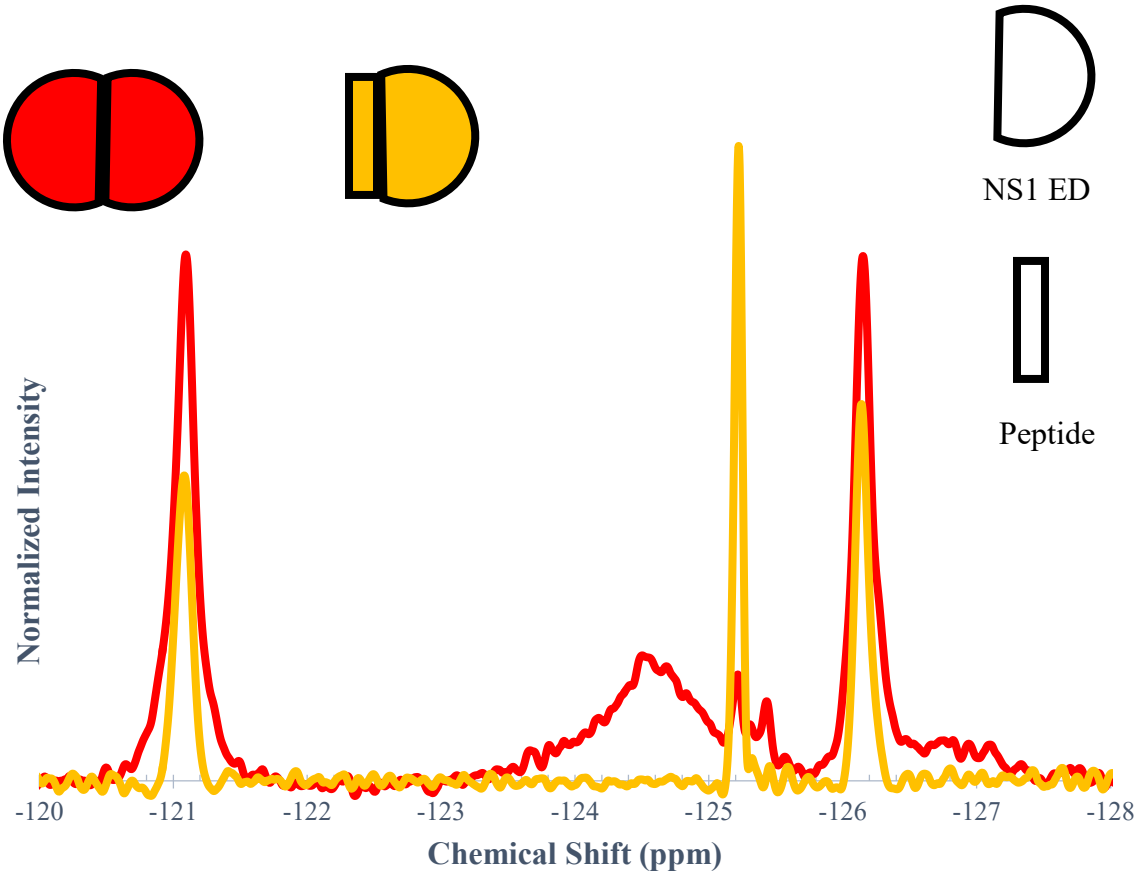


Figure 2-5 – Designed peptide binds near tryptophan 187 of the NS1 ED

The designed peptide significantly perturbs the NS1 effector domain. Unbound 200 μM 5-fluoro-tryptophan labeled ED (red) in a low-salt buffer exhibits the characteristic broad dimer peak while after the addition of peptide at 1:1 ratio (yellow) causes the broad peak to become an extremely intense narrow peak as observed upon F2F3 binding or ED in a high-salt buffer.

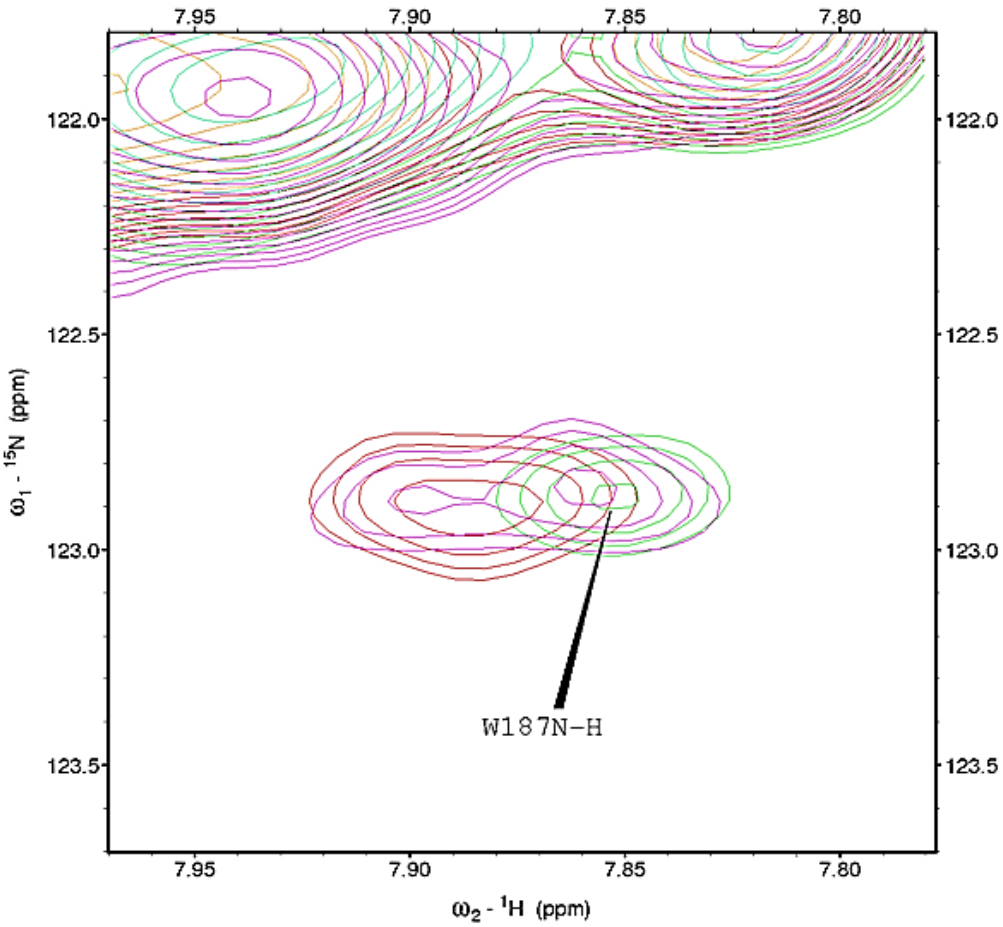


Figure 2-6 – Peptide induces a chemical shift of the W187 backbone

The peptide binding to W187 is validated by an NH-HSQC experiment. Green – 100 µM ED, Purple – 100 µM ED + 100 µM peptide, Red – 100 µM ED + 150 µM peptide.

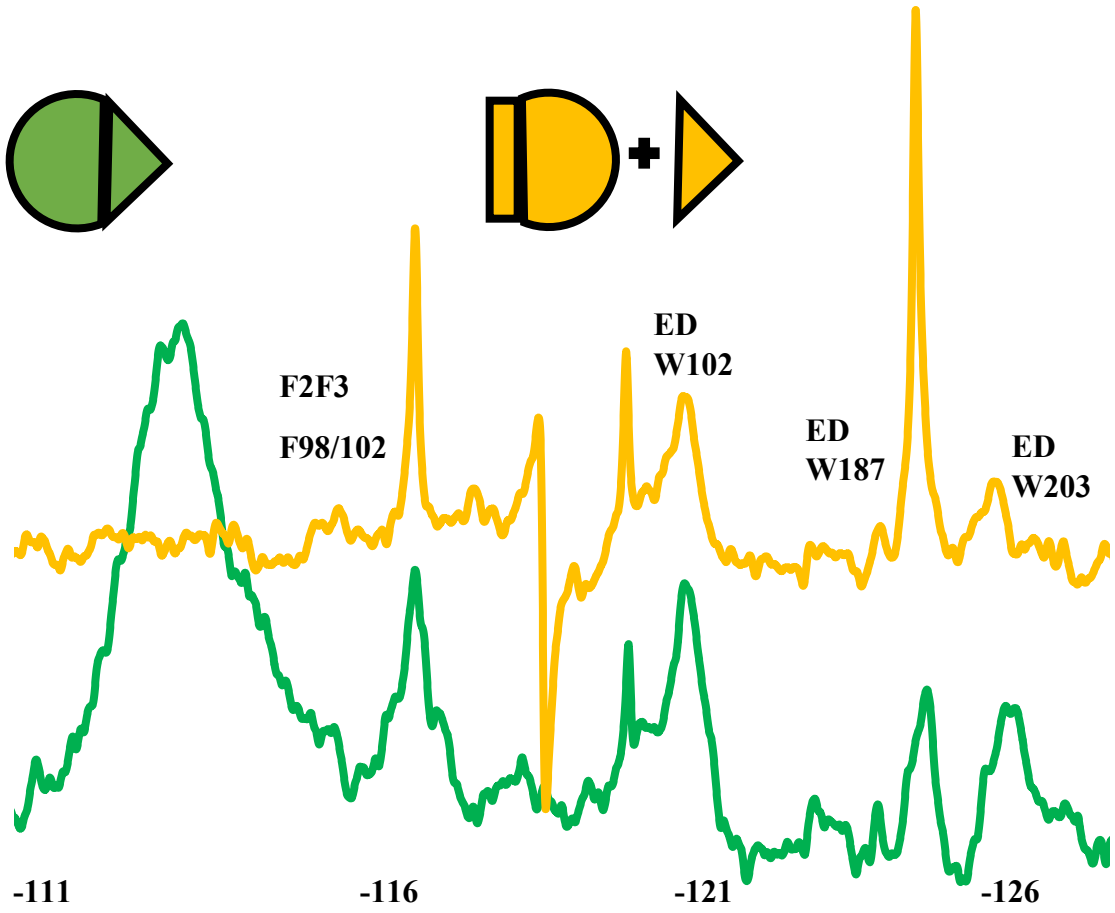


Figure 2-7 – Designed peptide inhibits CPSF30-F2F3 binding to NS1A ED

In the presence of excess CPSF30-F2F3 in phosphate buffered saline, the designed peptide induces significant chemical shift perturbations of the F2F3 binding-responsive peaks in addition to the effector domain W187 peak. Considerable intensification of the W187 peak occurs as seen in experiments without the presence of F2F3. In addition, the new, broad peak from F2F3 observed upon binding to NS1A ED disappears and a new partially phased peak evolves.

ADDENDUM

This section relates to partially completed projects which are intended to be continued and published.

DESIGN OF AN INFLUENZA A NS1 RNA-BINDING DOMAIN INHIBITOR

Introduction

The non-structural 1 protein of influenza A contains a domain shown to bind dsRNA and is highly conserved in members of the influenza A family (Figure 3-1 – Double-stranded RNA binds the NS1 RNA-binding domain along a cleft formed at the homodimer interface). In addition to high conservation between various strains of influenza A high conservation with members of the influenza B family is also observed. Using our peptide design software, *ProtCAD*, we have designed short α-helix peptides which may bind the RNA-binding interface produced when NS1 RBD dimerizes.

The RNA-binding interface produced during homodimerization of NS1 features a very electropositive charge and significant numbers of arginine and lysine amino acids contributed by both mononers (Figure 3-2 – The RNA-binding domain homodimer is a suitable target for a short α helix). In both crystal and NMR conditions, the arginine amino acid 38 plays a critical role forming strong salt-bridge interactions with the negatively charged polyphosphate backbone of dsRNA. The binding of dsRNA is expected to be favorable with long-range opposite charge interaction recruiting the RNA

molecule to the binding interface and local short-range interactions satisfying the charge with greater energy than simply binding to free water.

We have designed peptides for interfaces with structural information of a protein-protein interaction; however, we have not attempted to design a peptide to bind a site without prior protein-protein interaction structural data. In this circumstance, the size of the binding cleft was within the realm of binding potential for an α helix and we determined that the pocket had potential to test our automated design system in *ProtCAD*.

Methods and discussion

Initial selection of the binding target was made using the 1NS1 PDB structure of the NS1 RBD. This structure was chosen due to the simplicity of starting from a structure without dsRNA already bound. An idealized 11 amino acid alpha helix was created using *ProtCAD* and populated with a D-asparagine N-terminal cap, alanine, followed by D-glutamine and glycine in positions 10 and 11, respectively. Alanine was chosen due to the high helix propensity as well as lack of polar, charge, or significant hydrophobic interaction (Figure 3-3 – Initial design of an idealized helix).

The final structure was manually docked and minimized in AMBER12 to remove clashes. Our *ProtCAD* tool, ProtEvolver, was used and amino acids 2-9 were randomized to random amino acids excluding glycine, proline, and cysteine and allowed to mutate *in silico* iteratively for at least 18 cycles per replicate with additional cycles for each energetically favorable mutation. All amino acids were allowed for the initial mutation

cycle except for glycine, proline, and cysteine. Glycine and proline were excluded due to their inherently low helix propensity and cysteine was prohibited in order to prevent unwanted disulfide bond formation.

An initial simulation which produced 150 candidate sequences was conducted. ProtEvolver was setup to save sequence strings followed by total system energy and binding energy. Output sequences were sorted and used to manually force ideal amino acids in certain locations as well as reduce the total pool of amino acids available for mutation (Figure 3-4 – Designed sequences with optimal binding scores are chosen for further evaluation). Positions 7 and 8 were fixed to be aspartic acid and alanine, respectively, for example. Amino acids arginine, asparagine, histidine, isoleucine, lysine, and valine were additionally excluded from all positions due to poor performance in the first round of mutations (Figure 3-5 – Iterative evolution cycle data was used to refine the potential sequence pool for evolution).

The 10 peptides in the bottom 25 lowest total energy and 25 lowest binding energy were examined using AMBER's MMPBSA.py tool. Simulations were run and binding energies were calculated and compared to the MMPBSA derived generalized born energy from a simulation of the PDB 2ZKO after minization. The target energy to beat was -7.0 kcal/mol. The lowest MMPBSA energy peptide from the first cycle was calculated to be -19.0 kcal/mol and the lowest from the second was -24.1 kcal/mol. The sequence of the best peptide is D-N WWFDYDAW D-Q G.

The MMPBSA derived energetically best peptide from both the first and second cycle were synthesized by Tufts University Peptide Core Facility and examined using circular dichroism and nuclear magnetic resonance. Peptides were added to 100% ^{15}N,

5% ^{13}C labeled NS1 RBD at 250 μM in pH 8.0 buffer [50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 2.5 (v/v) glycerol, 10 mM dithiothreitol (DTT)]. A standard 2D NH-HSQC experiment was run and chemical shifts between apoprotein NS1 RBD and NS1 RBD mixed with peptides were observed.

Chemical shift perturbations were observed with the addition of both peptides, from round 1 and 2, however peptides from round 1 induced relatively weak chemical shift perturbations compared to the peptide from round 2. Specifically, arginine 38 was observed as a single peak in the apoprotein experiment but was observed to be split and shifted in the bound state indicative of strong binding (Figure 3-6 – Designed peptide from second cycle induces significant chemical shift of arginine 38 and peak splitting and Figure 3-7 – Chemical shift perturbations suggest strong binding to the same binding site as dsRNA). Peak splitting is observed due to the non-symmetric nature of binding to two different R38 residues from each independent RBD monomer. This splitting is also seen when dsRNA binds the same homodimer. According to molecular dynamics simulation data as well as chemical shift data, we believe this is due to the formation of cation-pi and salt-bridge interactions (Figure 3-8 – Molecular dynamics simulation of peptide suggests a combination of cation-pi and salt-bridge interactions). We believe this may be due to the asymmetric binding mode of the peptide to the protein since the interface contains two arginine 38 residues—one each from each monomer.

A fluorescence polarization experiment was performed using unlabeled NS1 RBD bound to fluorescently labeled dsRNA. In this experiment, the peptide have the opportunity to knock the labeled dsRNA out of the binding pocket changing its rotational polarization. Unfortunately, the experimental data has been lost though it suggested that

the peptides actually increased the total size of the complex suggesting non-competitive binding or binding without blocking dsRNA binding.

Future work

We have improved the design of ProtEvolver as discussed in a later section of this addendum. Further development has provided significantly faster and more precise calculations with the ability to produce better sequences. Better peptides should be designed and again examined using NMR. Fluorescence polarization should be used to validate competitive binding and a structure of the peptide-RBD complex can be solved.

Figures

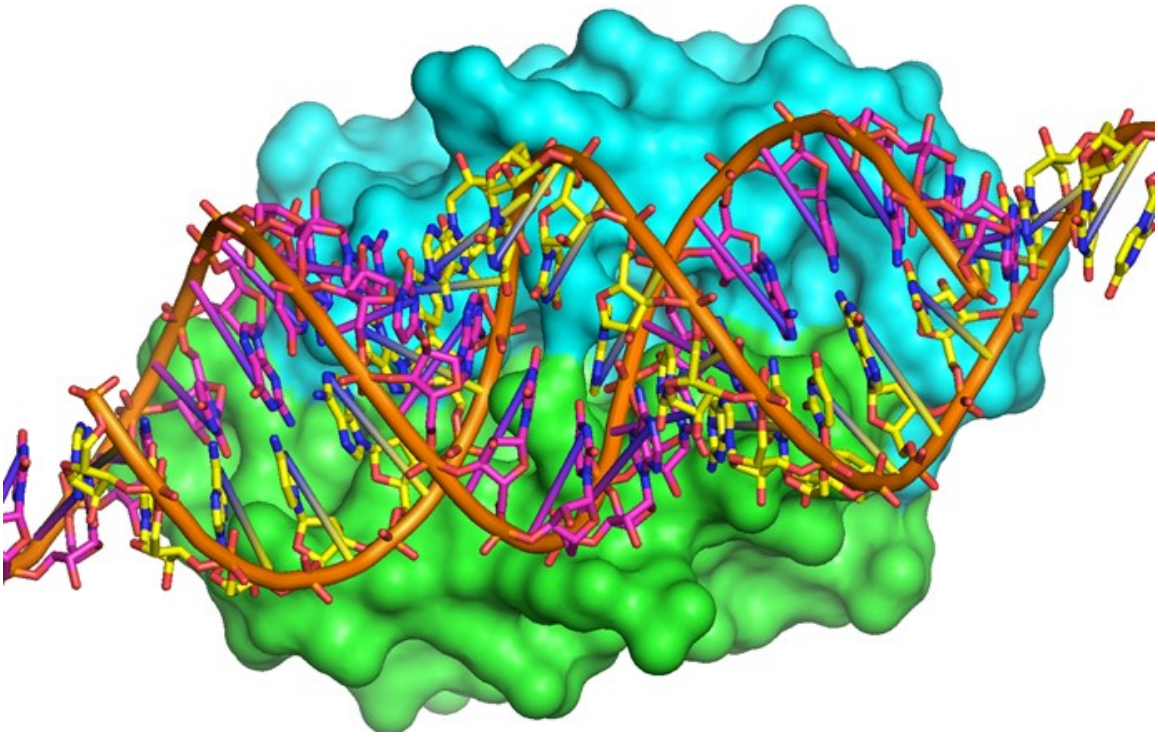


Figure 3-1 – Double-stranded RNA binds the NS1 RNA-binding domain along a cleft formed at the homodimer interface

The double stranded RNA (orange) binds a cleft formed by two NS1 RNA binding domain subunits (cyan and green) as they homodimerize. This cleft contains positively charged amino acids that intercalate between the polyphosphate backbone of the RNA molecule. PDB: 2ZKO

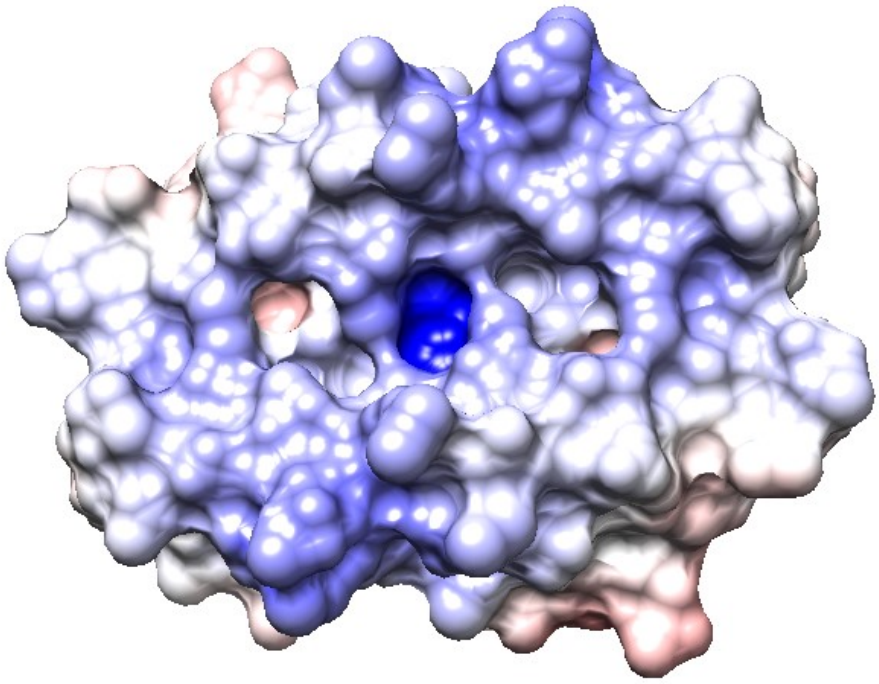


Figure 3-2 – The RNA-binding domain homodimer is a suitable target for a short α helix

The combination of positive charge as well as the presence of a large pocket with three deep clefts creates an interesting target for the design of a short helix peptide. PDB: 1NS1 colorized using the Adaptive Poisson Boltzmann Solver (APBS).

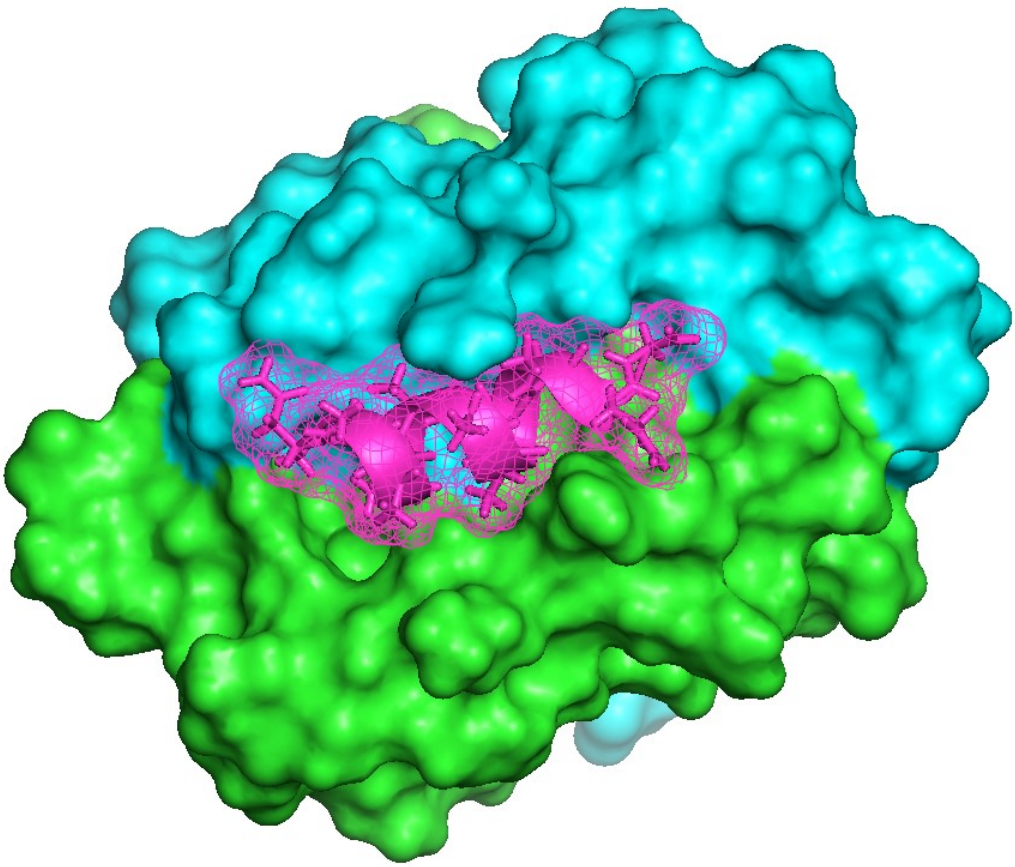


Figure 3-3 – Initial design of an idealized helix

An 11-amino acid α-helix containing a D-asparagine N-terminal cap, 8 alanine amino acids, a D-glutamine C-terminal cap followed by a glycine was equilibrated in AMBER molecular dynamics. This structure served as the starting structure for our *ProtCAD* simulations.

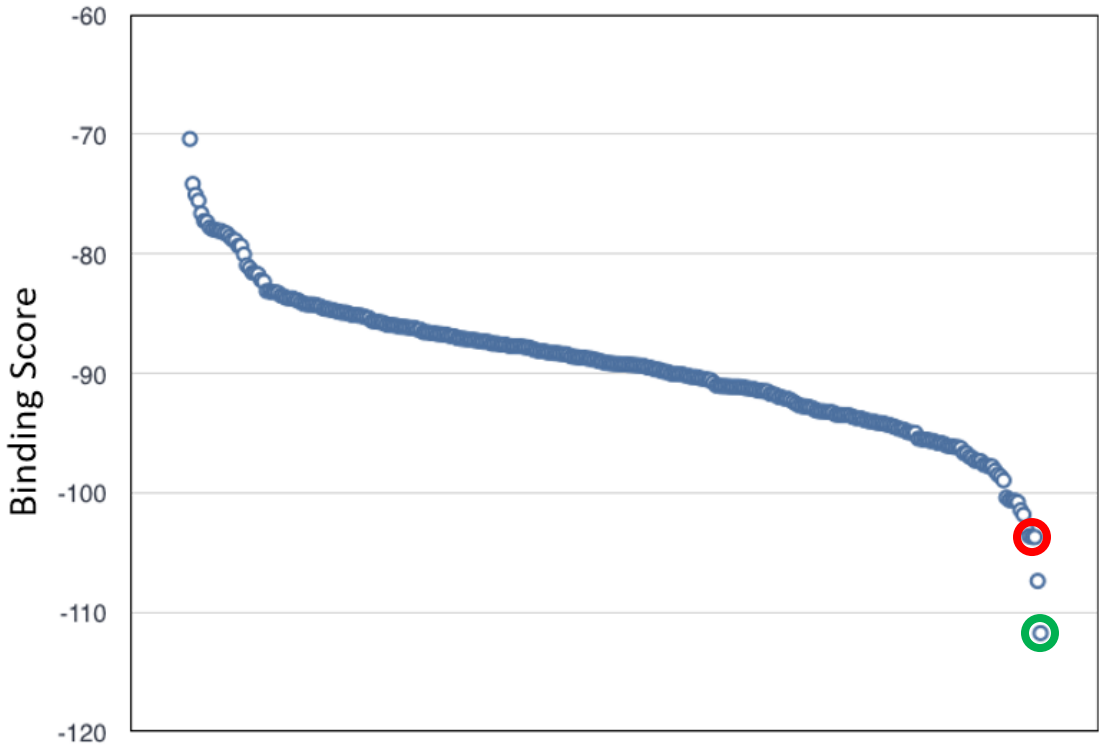


Figure 3-4 – Designed sequences with optimal binding scores are chosen for further evaluation

The top scores from two rounds of simulation were examined using the MMPBSA tool in AmberTools after binding simulation. The best sequence from the first round (red) which had a calculated binding energy of -19 kcal/mol and second round (green) with a calculated binding energy of -24 kcal/mol were chosen to be synthesized and examined using NMR.

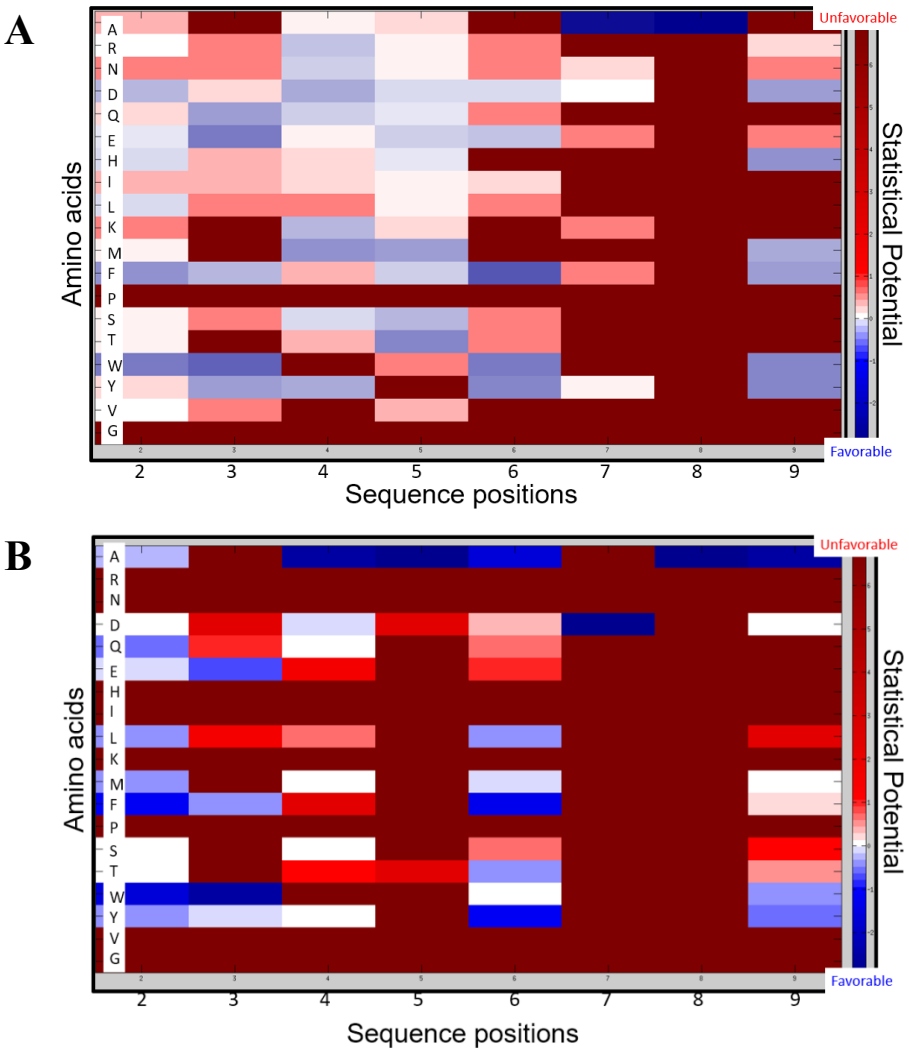


Figure 3-5 – Iterative evolution cycle data was used to refine the potential sequence pool for evolution

Amino acids were excluded depending on propensity from the first evolution cycle. Certain amino acids were chosen to be locked in for the second cycle based upon data from the first. The statistical potential was derived from the observed sequence count per position over the course of 150 simulations. Frequently observed mutations are colored dark blue and rarely observed mutations are colored dark red.

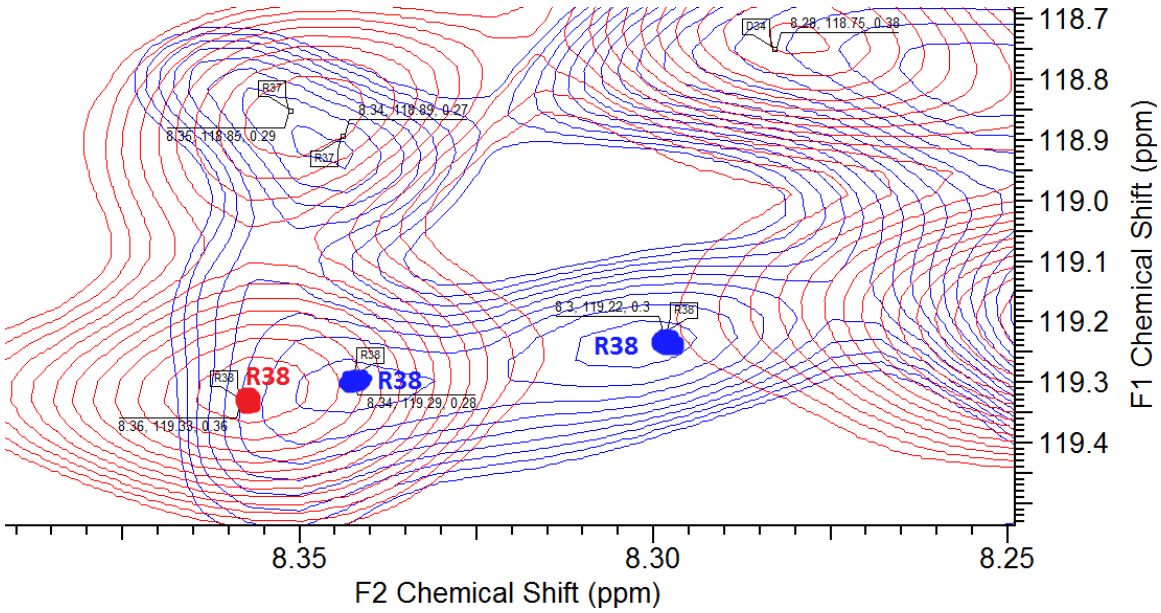


Figure 3-6 – Designed peptide from second cycle induces significant chemical shift of arginine 38 and peak splitting

Arginine 38 shifts and splits when in the presence of designed peptide. Red – 100 µM NS1 RNA binding domain. Blue – 100 µM NS1 RNA binding domain and 100 µM Peptide.

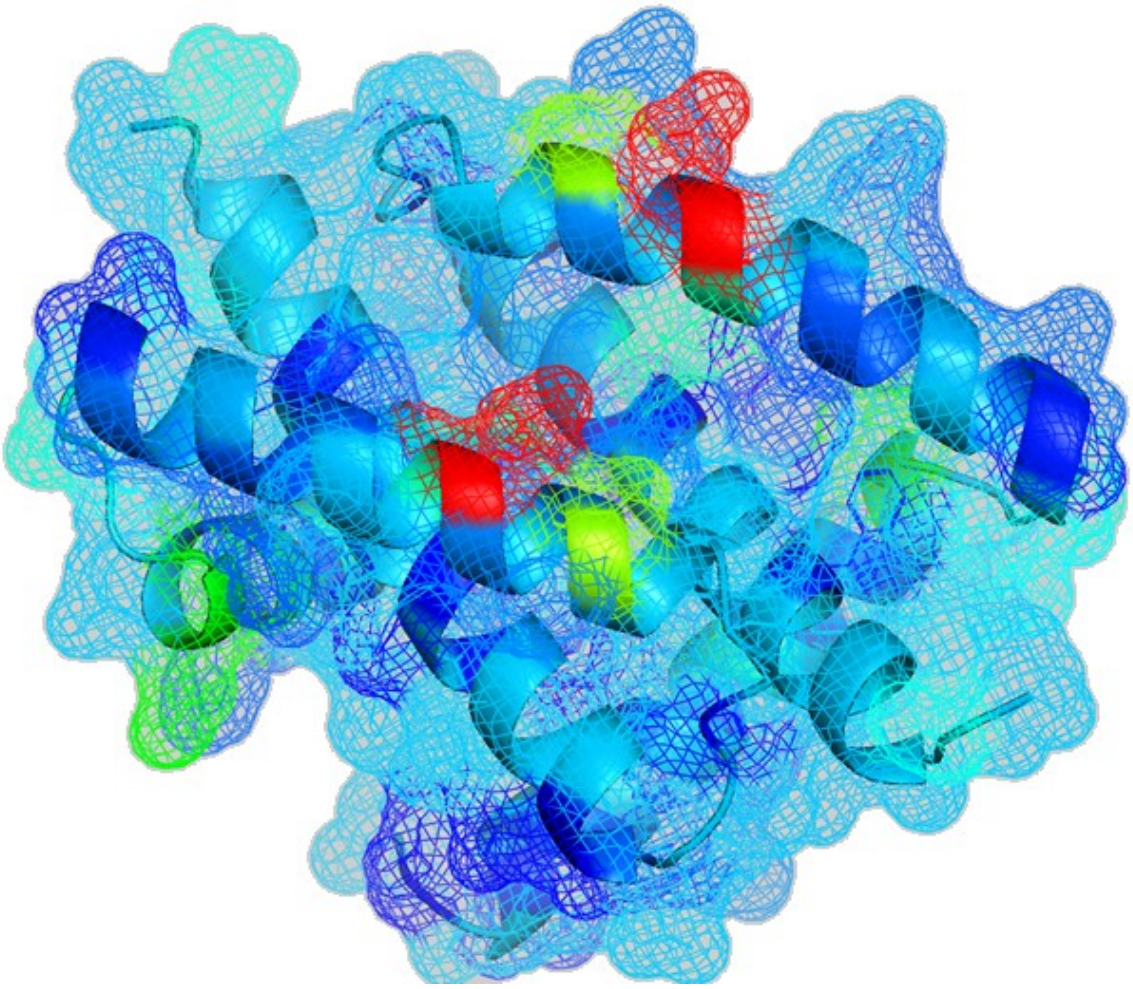


Figure 3-7 – Chemical shift perturbations suggest strong binding to the same binding site as dsRNA

Chemical shift perturbations are strongest at the arginine 38 amino acid with moderate CSP observed at hinge residues in the interfaces between helices of the RBD. Red – significant CSP, green – moderate CSP, dark and light blue – low CSP.

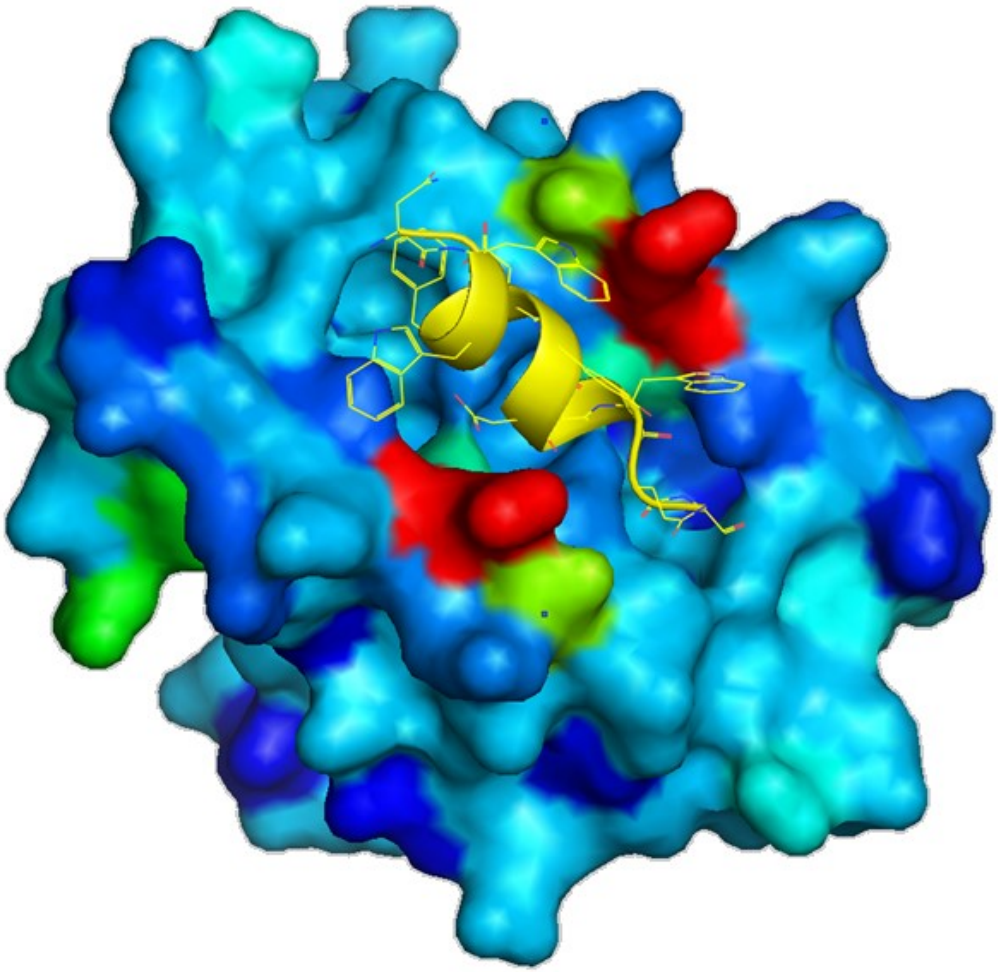


Figure 3-8 – Molecular dynamics simulation of peptide suggests a combination of cation-pi and salt-bridge interactions

AMBER simulation suggests a strong cation-pi interaction between a tryptophan on the peptide and arginine 38 helps to stabilize the interaction. In addition, an aspartic acid-arginine 38 interaction on the other RBD subunit helps to anchor and further stabilize the interaction.

CONTROL OF PEPTIDE AGGREGATION AND HIGHER ORDER STRUCTURE FORMATION VIA STEREOCHEMISTRY

Introduction

In the process of designing peptides for the NS1 Effector Domain and RNA Binding Domain, control peptides were synthesized with N- and C-terminal caps in L-conformation or glycine. We expected this to allow us to examine impact of our D-amino acid caps on helicity. Unfortunately, many of our control peptides were highly insoluble and formed thick precipitates and hydrogels. While this was not our intention, we were interested in the impact of D-amino acids on short peptide stability and decided to further analyze the control peptides.

D-amino acids have been observed in various neurodegenerative diseases such as Parkinson's disease and Alzheimer's disease and may play a role in protein aggregation and misfolding.[11,12] Besides potentially contributing to disease states, amyloids have been proposed to be used for delivery of drug compounds or viruses to target tissues.[106] By engineering amyloids using D-amino acids, it may be possible to design carriers for hard-to-deliver compounds and the understanding of chirality on protein folding may allow for better therapies to prevent degenerative diseases.

Methods and discussion

Peptides were derived from the NS1 ED inhibitor work. The initial designed ED peptide sequence is D-N Y F Y S L F D-Q G. Various substituted peptides were created:

Pep2	D-N	Y	F	Y	S	L	F	D-Q	G
Pep2NdQ	N	Y	F	Y	S	L	F	D-Q	G
Pep2dNQ	D-N	Y	F	Y	S	L	F	Q	G
Pep2NQ	N	Y	F	Y	S	L	F	Q	G
Pep2GdQ	G	Y	F	Y	S	L	F	D-Q	G
Pep2dNG	D-N	Y	F	Y	S	L	F	G	G
Pep2GG	G	Y	F	Y	S	L	F	G	G
Pep2A	D-N	Y	F	A	S	L	F	D-Q	G
Pep2K	D-N	Y	F	Y	S	K	F	D-Q	G
Pep2AK	D-N	Y	F	A	S	K	F	D-Q	G

These peptides were created in order to examine the effect of L vs D-amino acid as well as allowing glycine, an amino acid which has no chirality, to be examined.

In all cases, solubility was examined by mixing the lyophilized peptides with pH 7.4 phosphate buffered saline (PBS) to target a final concentration of 0.5 mM of peptide. Prepared peptide samples were centrifuged at 13,000 rpm for 2 minutes to remove any precipitate and peptide concentrations were determined on a UV/VIS spectrophotometer. In addition to solubility measurements, transmission electron micrographs were collected for each peptide, dynamic light scattering was performed to measure peptide aggregate dimensions, as well as a thioflavin T assay to determine whether peptides formed amyloids.

In the case of the three variants with a glycine substitution for a D-amino acid cap, final peptide concentrations were too low to be observed by UV/VIS. The starting peptide with two D-amino acid caps was the most soluble with 0.478 mM concentration compared to the predicted 0.500 mM by the initial mass. Pep2NdQ followed with 0.220 mM final concentration, Pep2NQ with 0.085 mM final concentration, and finally dNQ with 0.015 mM final concentration. Peptides Pep2A, Pep2K, and Pep2AK were all completely soluble with 0.5 mM measured final concentration and no precipitate formation.

Peptide samples were examined in the same conditions using dynamic light scattering (DLS) performed on a Malvern Zetasizer ZS unit. Measured size radii using a standard spherical model range from 0.3 to 600 nm depending on the peptide (Figure 3-9 – Dual D-amino acid caps decreases higher-order structure formation). Pep2AK, Pep2A, and Pep2K all exhibited average sizes of 0.3 to 0.5 nm while Pep2 averaged 0.8 nm. Pep2NdQ exhibited multiple size states with 0.6 nm, 40 nm, 600 nm, and 1200 nm with 0.6 nm and 600 nm being the predominant species. Pep2NQ was roughly 35 nm in radius and PepdNQ was approximately 100 nm. Glycine variants were too insoluble to be detected by DLS. Variants of Pep2 with less hydrophobic sequences appeared to be monomers in solution with no aggregation. While the spherical model is not the optimal model to examine these peptides, it served as a reasonable comparative baseline as several peptides appeared to aggregate as linear amyloid composites with no observed globular structure.

Transmission electron micrographs were collected using osmium tetroxide negative stain using a JEOL 1200EX electron microscope with peptides bound to a

copper grid. Several representative images were recorded. Three primary phenotypes were observed: left-handed twisted fibers, planar rods or sheets, and curved fibers. All peptides besides Pep2A, Pep2K, and Pep2AK exhibited fiber formation in TEM while Pep2 only rarely exhibited this phenomenon compared to the other peptides (Figure 3-11 – Pep2GdQ, Pep2dNQ, and Pep2NQ produce planer rods or sheets). Twisted fibers were observed in Pep2, Pep2NdQ and Pep2dNG samples, rods or sheets were seen in Pep2dNQ, Pep2NQ, and Pep2GdQ (Figure 3-10 – Pep2, Pep2NdQ, and Pep2dNG produce twisted fibers). Pep2GG exhibited a curved fiber phenotype (Figure 3-12 – Pep2GG produces curved fibers).

A thioflavin T binding assay to determine cross-β amyloid character was performed in which the dye binds beta sheet-rich structures inducing enhanced fluorescence and a red shift of emission spectra (Figure 3-13 – Pep2 shows the least amyloid-like signal in a thioflavin T staining assay). Pep2A, Pep2K, and Pep2AK were not examined as they did not exhibit amyloid character in electron micrographs. Pep2 exhibited the weakest Thioflavin T binding by a large margin compared to the other peptides.

Molecular dynamics were run on single Pep2, Pep2NdQ, Pep2dNQ and Pep2NQ peptides for 200 nanoseconds at 300K to examine backbone hydrogen bonding and α-helicity. Hydrogen bonding propensities were calculated by observing i carbonyl to i+4 amide hydrogen bonds (Figure 3-14 – Simulation predicts Pep2 to have significantly more α-helix character than peptides with only one or no D-amino acid caps). Pep2 formed significantly more hydrogen bonds within the backbone compared to Pep2NQ and Pep2NdQ followed by the least helical Pep2dNQ.

Future work

Modeling of the peptides should be completed as well as molecular dynamics and helicity determination of the remaining peptides. A repeat of thioflavin T measurements may be useful along with data for the Pep2A, Pep2K, and Pep2AK.

Figures

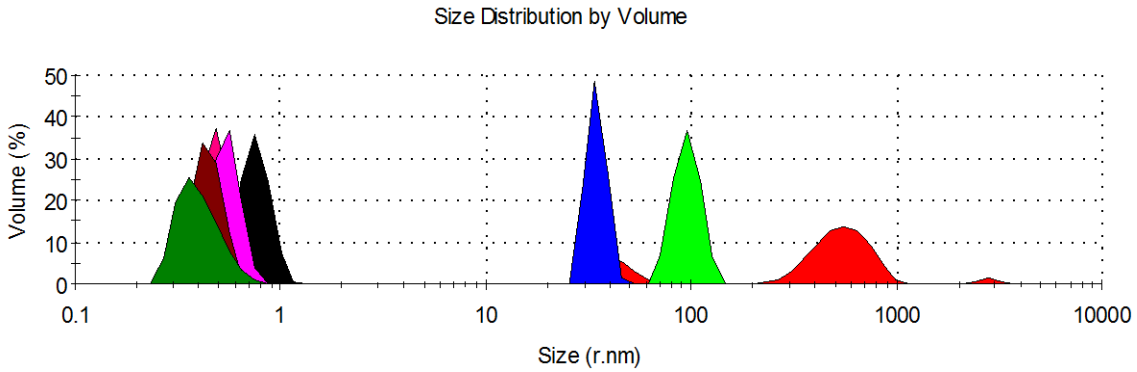


Figure 3-9 – Dual D-amino acid caps decreases higher-order structure formation

Pep2AK (olive), Pep2K (maroon), Pep2A (magenta), and Pep2 (black) are all predominantly monomer in size with sub-nanometer radii. Pep2NQ (blue) is followed by Pep2dNQ (light green), and Pep2NdQ (red). The particle size of Pep2NQ is over an order of magnitude larger than the peptides with two D-amino acid caps. Pep2dNQ is two orders of magnitude larger as is Pep2NdQ. The lack of D-amino acid caps on the N and C-terminus greatly reduces the ability for the peptides to stay soluble and not form amyloid fibers.

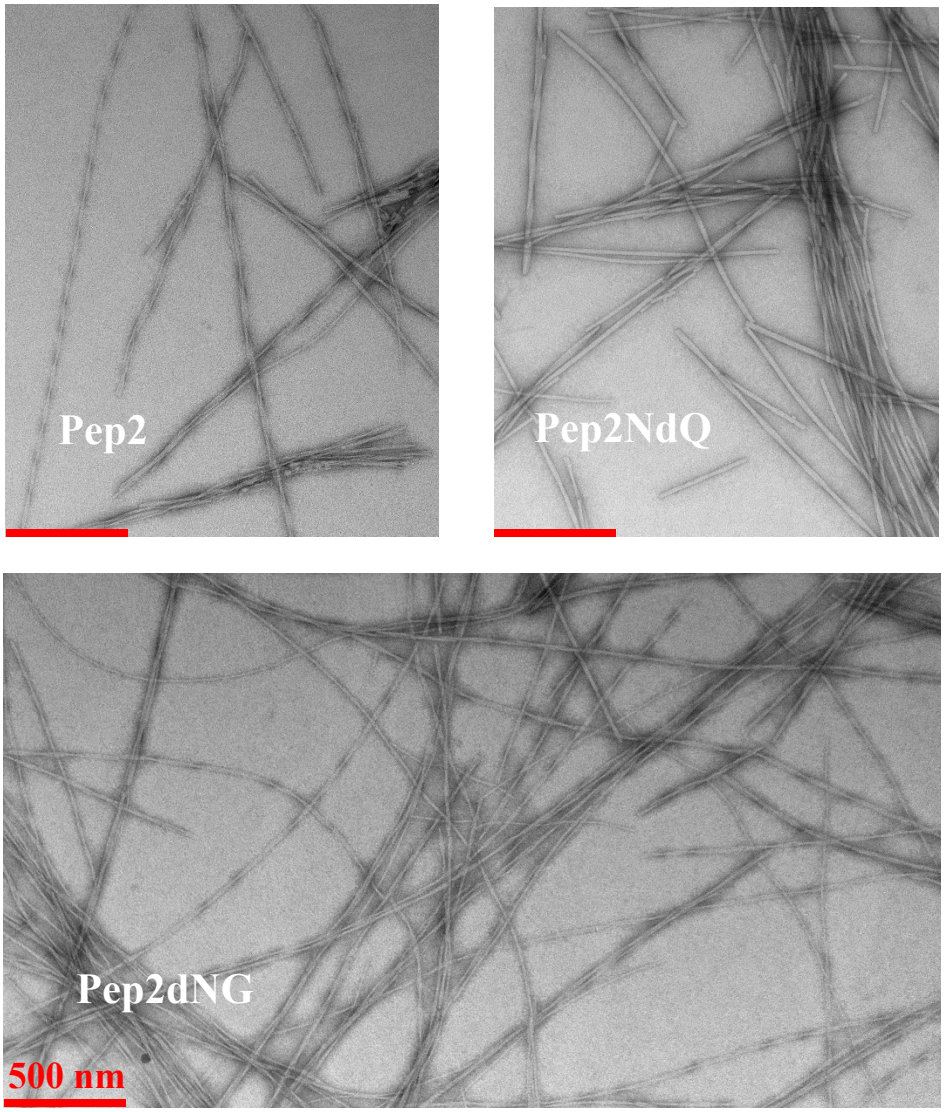


Figure 3-10 – Pep2, Pep2NdQ, and Pep2dNG produce twisted fibers

The presence of L or D-asparagine in the first position with a C-terminal XG where X is an amino acid compatible with a D-conformation forms a twisted fiber.

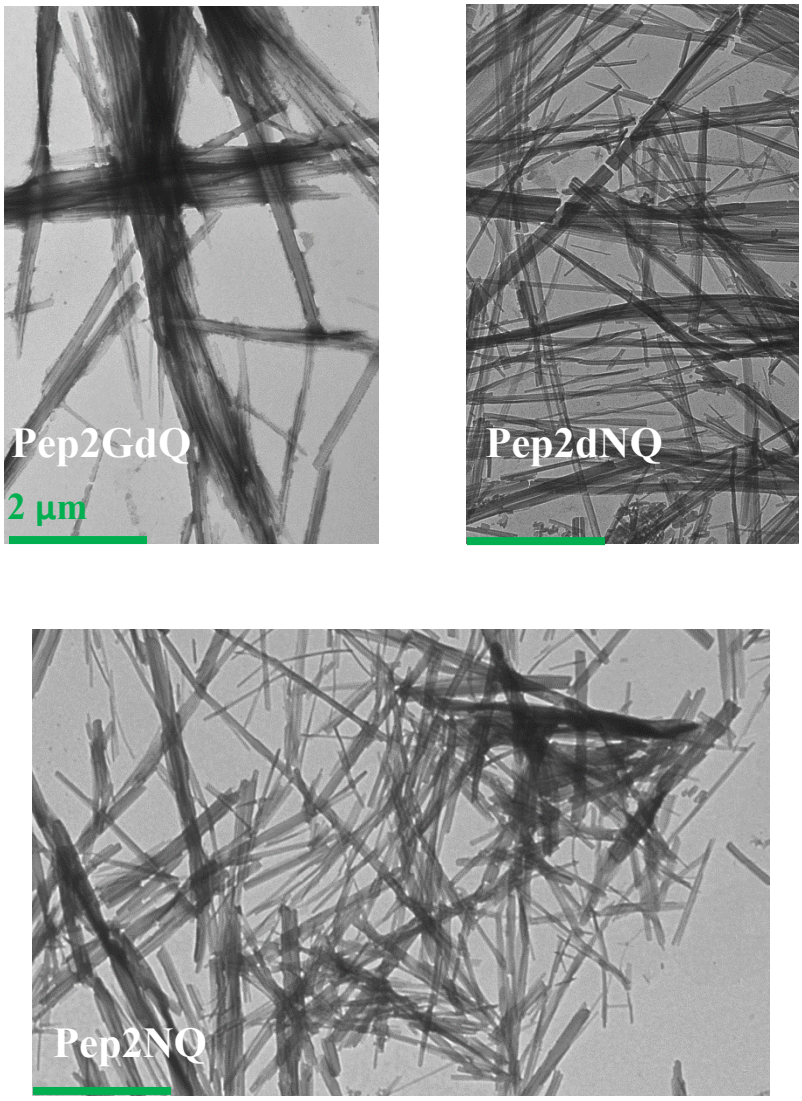


Figure 3-11 – Pep2GdQ, Pep2dNQ, and Pep2NQ produce planer rods or sheets

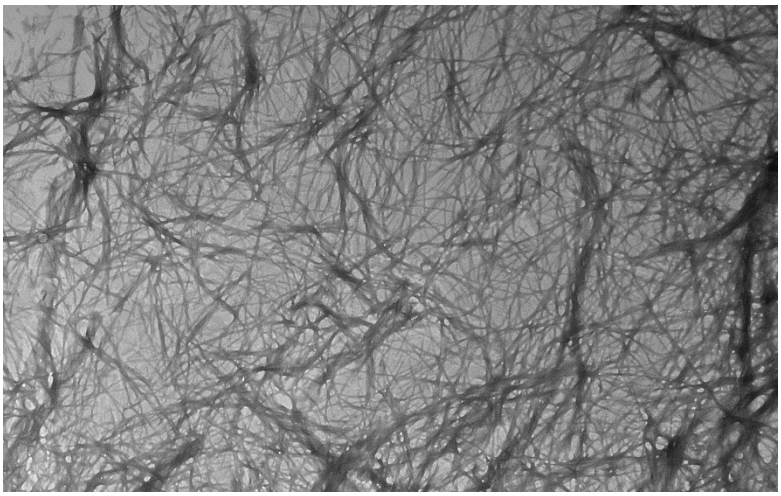


Figure 3-12 – Pep2GG produces curved fibers

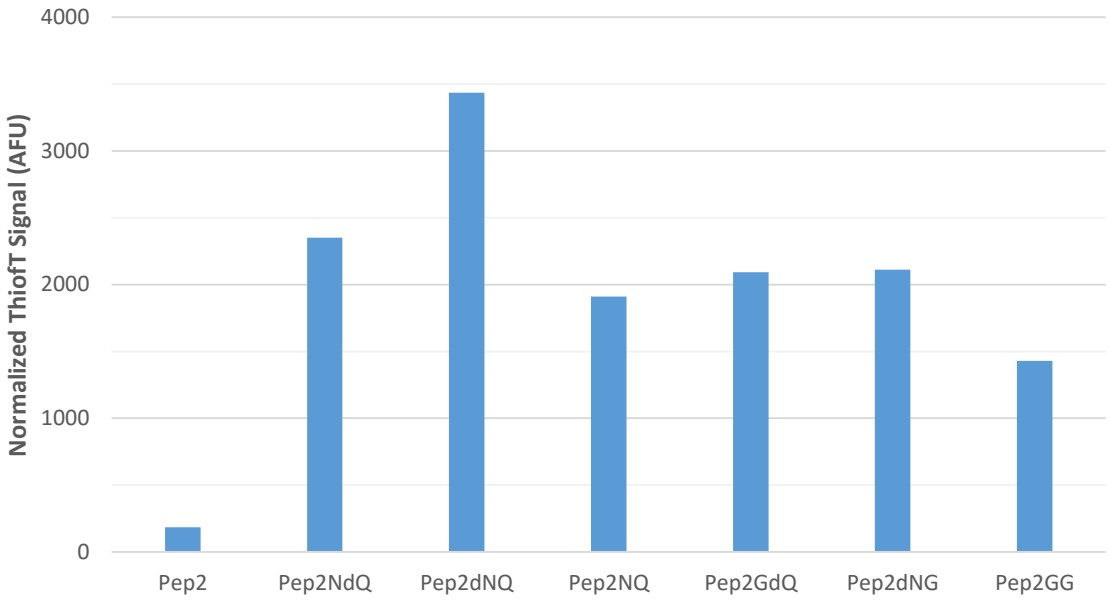
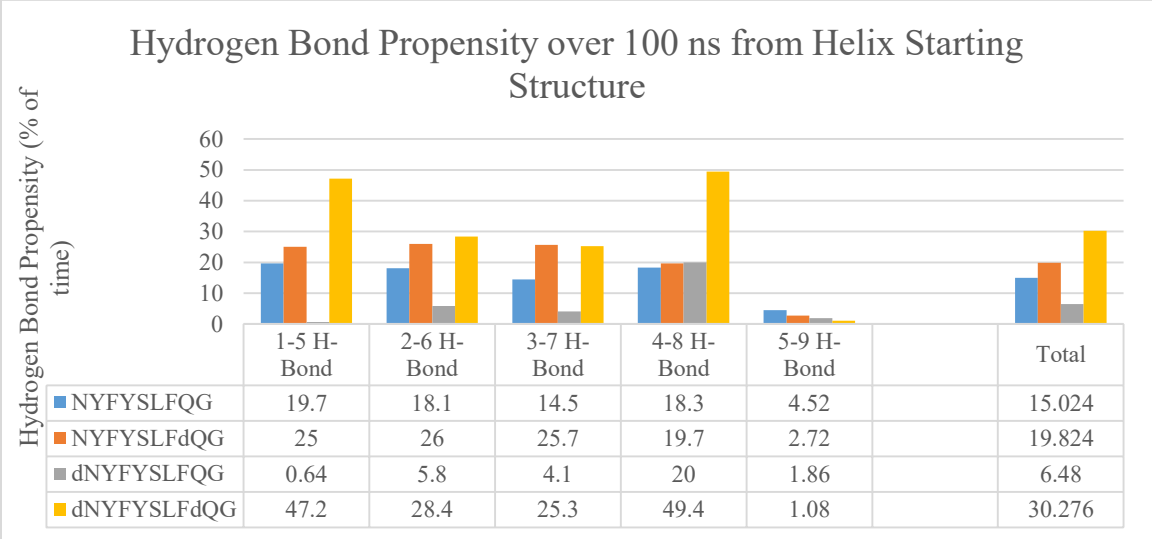


Figure 3-13 – Pep2 shows the least amyloid-like signal in a thioflavin T staining assay

Peptides with one or no D-amino acid caps bind thioflavin T considerably more than Pep2 which contains two D-amino acid caps. All peptides besides Pep2 appear to be almost all amyloid.



	1-5 H-Bond	2-6 H-Bond	3-7 H-Bond	4-8 H-Bond	5-9 H-Bond		Total
NYFYSLFQG	19.7	18.1	14.5	18.3	4.52		15.024
NYFYSLFdQG	25	26	25.7	19.7	2.72		19.824
dNYFYSLFQG	0.64	5.8	4.1	20	1.86		6.48
dNYFYSLFdQG	47.2	28.4	25.3	49.4	1.08		30.276

Figure 3-14 – Simulation predicts Pep2 to have significantly more α-helix character than peptides with only one or no D-amino acid caps

By observing the propensity of backbone hydrogen bonds between i to i+4 amino acids during a molecular dynamics simulation, Pep2 (dNYFYSLFdQG) possesses hydrogen bonds between backbone carbonyl and amides considerably more than variants with fewer D-amino acids.

DESIGN OF A BYZANTINE FAULT TOLERANT DISTRIBUTED PROTEIN DESIGN SYSTEM

Introduction

Protein and peptide design software has been around for decades however distributed systems are not well documented. The Genome@Home project based out of Stanford University has been designing new sequences for the SH3 domain.[72] It relies on a central server coordinating smaller simulations on user computers. Rosetta, a very popular protein design software tool, is able to run on multiple cores on a local system, however a distributed implementation has not been created.

In the case of Genome@Home, if the central server fails, the design halts and contributing nodes are not able to progress the simulation. In addition, nodes disconnected from the central server are useless. The system relies on a central link passing information to and from the computing nodes.

We have designed a protein design tool that can run locally and can also run using a distributed method where sparse information is shared. Total design throughput is derived as a linear combination of the total effort put forth by all contributing nodes with no significant slowdown when many nodes are working to solve the problem.

Methods and discussion

Our technique involves using a cloud or network based filesharing system. In our case, we used Google Drive to communicate between the nodes however Dropbox, OneDrive, or even standard IP based networking protocols can be used. Each node

requires a ProtEvolver binary compiled with the same evolution options and the starting PDB file.

The system creates a file, in our case, finalsequences.out. This file is shared among all systems via the filesharing system. The file is constructed where each line in the file is a suitable final protein primary sequence. Each local node can calculate the total and binding energies of the sequence by mutating the staring PDB file with each sequence and determining the energies. Since structural data is not saved and energies are locally calculated, each local node can determine the fate of each sequence submitted to it from every node.

The ProtEvolver design software relies on a genetic algorithm that enhances "good sequence" probability during the mutation process in a linear manner starting from approximately the 500^{th} sequence, scaling until approximately 1500 sequences in which all mutations are derived from the sequence database. The lack of forced reliance of data from any particular node and spares total information allows each local node to error-check the others and discard potentially bad information. This attribute makes the design system immune to infectious or non-present nodes and increase the reliability of the system.

Future work

Further improvements to the protein design algorithm will help make it more efficient. A cryptographic hashing function of prior designed sequences could create a protein design blockchain allowing for higher security.

Figures

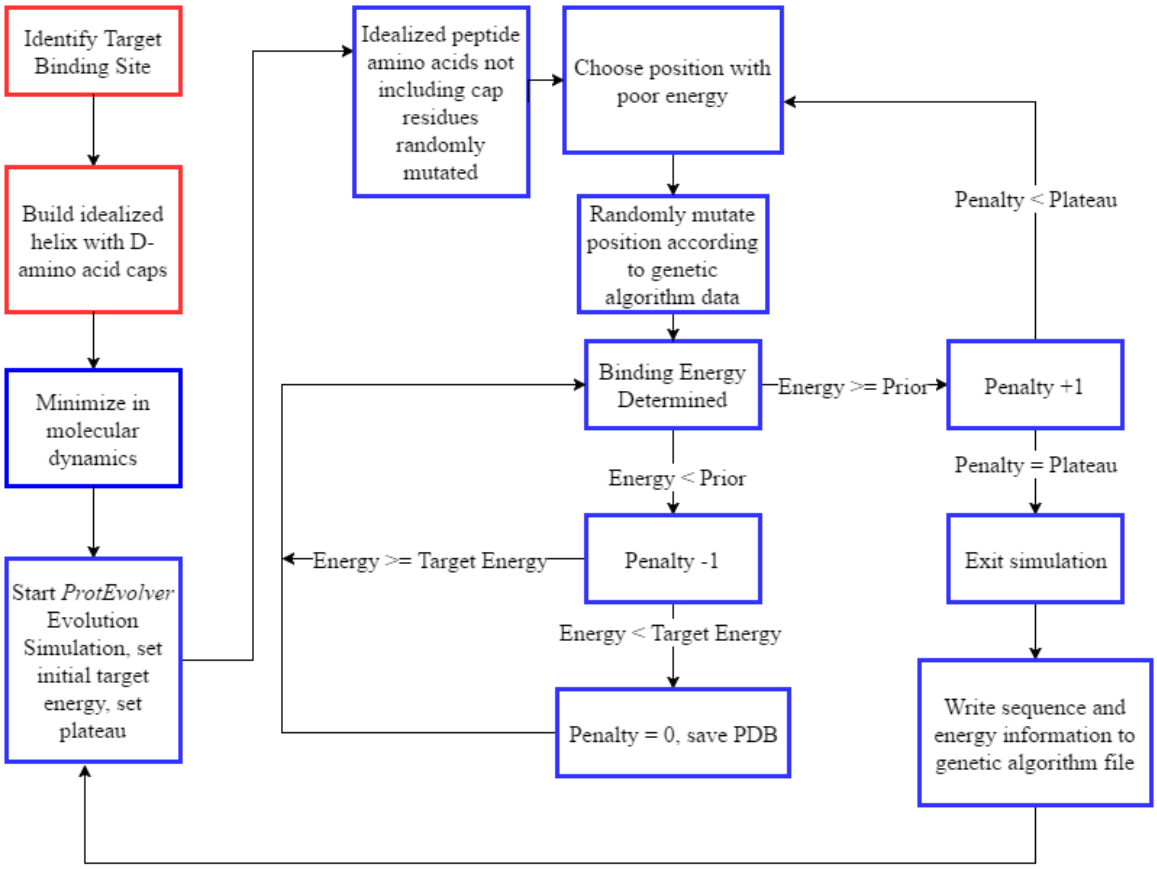


Figure 3-15 – Flow Chart of inhibitor design in *ProtEvolver*

APPENDIX

Trp-cage D-threonine Peak List:

```
# Number of dimensions 2
   1  7.440  1.093  1 U 2.91e+006        0 e 0   57   60
   2  7.859  7.420  1 U 2.55e+006        0 e 0   53   57
   3  8.192  7.419  1 U 5.66e+006        0 e 0   61   57
   4  4.302  7.422  1 U 1.98e+006        0 e 0   58   57
   5  1.112  7.419  1 U 3.35e+006        0 e 0   60   57
   6  7.441  4.280  1 U 6.45e+005        0 e 0   57   58
   7  8.140  4.293  1 U 9.31e+005        0 e 0   68   69
   8  8.047  4.291  1 U 9.17e+005        0 e 0   72   69
   9  3.921  4.298  1 U 1.65e+006        0 e 0   70   69
  10  3.715  4.297  1 U 1.74e+006        0 e 0   71   69
  11  8.304  7.840  1 U 4.04e+006        0 e 0   48   53
  12  7.439  7.840  1 U 2.84e+006        0 e 0   57   57
  13  2.877  7.839  1 U 3.48e+006        0 e 0   55   53
  14  3.049  7.840  1 U 1.28e+006        0 e 0   56   53
  15  7.439  8.171  1 U 5.54e+006        0 e 0   57   61
  16  4.300  8.080  1 U -4.54e+004        0 e 0   69   72
  17  4.172  8.096  1 U -9.74e+003        0 e 0   74   72
  18  3.877  8.095  1 U 2.15e+004        0 e 0   73   72
  19  7.439  7.419  1 U 9.18e+007        0 e 0   57   57
  20  8.140  8.028  1 U 2.75e+006        0 e 0   68   72
  21  8.048  8.030  1 U 2.76e+008        0 e 0   72   72
  22  8.025  8.255  1 U 3.61e+004        0 e 0   75   72
  23  4.163  3.848  1 U 2.68e+006        0 e 0   74   73
  24  3.860  4.175  1 U 6.68e+006        0 e 0   73   74
  25  8.907  8.452  1 U 2.33e+006        0 e 0    7   13
  26  8.473  8.887  1 U 2.34e+006        0 e 0   13    7
  27  8.907  8.886  1 U 1.12e+008        0 e 0    7    7
  28  4.356  8.886  1 U 6.89e+006        0 e 0    1    7
  29  4.267  8.888  1 U 1.79e+006        0 e 0    8    7
  30  4.268  8.453  1 U 2.33e+006        0 e 0    8   13
  31  4.188  8.451  1 U 2.48e+006        0 e 0   14   13
  32  8.093  8.453  1 U 3.53e+006        0 e 0   18   13
  33  3.174  8.886  1 U 1.65e+006        0 e 0    6    7
  34  3.098  8.452  1 U 6.32e+006        0 e 0   15   13
```

```
35  1.737  8.886  1 U 6.23e+006       0 e 0    9    7
36  1.519  8.886  1 U  9.2e+005       0 e 0   10    7
37  1.737  8.452  1 U 1.42e+006       0 e 0    9   13
38  1.517  8.453  1 U 6.68e+005       0 e 0   10   13
39  0.957  8.886  1 U 4.74e+005       0 e 0   11    7
40  0.909  8.886  1 U 5.55e+005       0 e 0   12    7
41  0.912  8.453  1 U 4.07e+005       0 e 0   12   13
42  1.948  0.895  1 U  5.1e+006       0 e 0   20   24
43  8.094  8.073  1 U 1.57e+008       0 e 0   18   18
44  8.473  8.073  1 U 3.74e+006       0 e 0   13   18
45  7.852  8.074  1 U 3.33e+006       0 e 0   25   18
46  8.050  7.832  1 U 4.13e+006       0 e 0   31   25
47  8.094  7.833  1 U  3.1e+006       0 e 0   18   25
48  7.855  7.835  1 U 2.13e+008       0 e 0   25   25
49  7.853  8.029  1 U 4.13e+006       0 e 0   25   31
50  8.473  8.452  1 U 1.43e+008       0 e 0   13   13
51  8.340  8.032  1 U 3.78e+006       0 e 0   41   31
52  3.829  7.837  1 U 1.78e+006       0 e 0   49   53
53  4.272  7.831  1 U 6.21e+005       0 e 0    8   25
54  3.993  7.835  1 U 3.81e+006       0 e 0   26   25
55  3.830  7.832  1 U  2.6e+006       0 e 0   19   25
56  3.643  3.054  1 U 1.61e+007       0 e 0   63   64
57  4.299  3.177  1 U 3.21e+006       0 e 0   32   33
58  8.050  3.177  1 U 3.63e+006       0 e 0   31   33
59  8.050  3.521  1 U 4.14e+006       0 e 0   31   34
60  4.299  3.535  1 U 1.66e+006       0 e 0   32   34
61  3.542  3.170  1 U 1.43e+007       0 e 0   34   33
62  3.199  3.524  1 U 1.27e+007       0 e 0   33   34
63  3.074  3.622  1 U 1.44e+007       0 e 0   64   63
64  4.187  3.522  1 U 1.74e+006       0 e 0   14   34
65  4.187  3.172  1 U 5.79e+005       0 e 0   14   33
66  3.543  4.271  1 U 1.53e+006       0 e 0   34   32
67  3.198  4.276  1 U 1.64e+006       0 e 0   33   32
68  8.049  4.275  1 U 1.05e+006       0 e 0   31   32
69  8.049  8.029  1 U 2.76e+008       0 e 0   31   31
70  8.050  8.319  1 U 3.35e+006       0 e 0   31   41
71  7.221  8.320  1 U 1.72e+006       0 e 0   37   41
72  8.301  8.319  1 U 3.24e+006       0 e 0   48   41
```

```
 73  8.304  8.284  1 U 1.54e+008       0 e 0   48   48
 74  7.859  8.284  1 U 3.69e+006       0 e 0   53   48
 75  8.340  8.319  1 U 1.32e+008       0 e 0   41   41
 76  3.166  4.347  1 U 6.45e+005       0 e 0    2    1
 77  3.045  4.350  1 U 6.08e+005       0 e 0    3    1
 78  8.907  4.334  1 U 2.03e+006       0 e 0    7    1
 79  8.901  4.263  1 U 2.39e+005       0 e 0    7    8
 80  8.473  4.247  1 U  8.5e+005       0 e 0   13    8
 81  8.472  4.168  1 U 1.27e+006       0 e 0   13   14
 82  8.069  4.212  1 U 2.84e+004       0 e 0   18   14
 83  7.029  4.167  1 U 9.26e+005       0 e 0   16   14
 84  9.707  9.687  1 U 4.15e+007       0 e 0   36   36
 85  7.114  9.687  1 U 2.38e+006       0 e 0   35   36
 86  7.304  9.687  1 U 9.24e+005       0 e 0   39   36
 87  9.764  7.303  1 U 1.41e+005       0 e 0   36   39
 88  9.764  7.118  1 U -4.6e+003       0 e 0   36   35
 89  8.339  7.203  1 U 2.45e+006       0 e 0   41   37
 90  4.300  7.095  1 U 1.91e+006       0 e 0   32   35
 91  3.198  7.095  1 U 2.88e+006       0 e 0   33   35
 92  3.541  7.095  1 U 1.31e+006       0 e 0   34   35
 93  7.828  7.807  1 U 1.39e+008       0 e 0    4    4
 94  7.828  7.191  1 U 5.09e+007       0 e 0    4    5
 95  7.212  7.807  1 U  5.2e+007       0 e 0    5    4
 96  7.217  7.196  1 U  3.8e+008       0 e 0   38   38
 97  7.476  7.455  1 U 8.97e+007       0 e 0   27   27
 98  7.476  6.934  1 U 4.08e+007       0 e 0   27   28
 99  6.955  7.456  1 U 4.48e+007       0 e 0   28   27
100  6.955  6.935  1 U 9.08e+007       0 e 0   28   28
101  3.173  7.808  1 U 1.57e+006       0 e 0    2    4
102  3.047  7.808  1 U 1.07e+006       0 e 0    3    4
103  2.408  7.456  1 U 7.68e+005       0 e 0   30   27
104  7.303  7.284  1 U 1.55e+008       0 e 0   39   39
105  7.120  7.119  1 U 6.29e+006       0 e 0   35   35
106  7.219  7.283  1 U 1.77e+007       0 e 0   38   39
107  7.217  7.093  1 U 1.99e+007       0 e 0   38   40
108  7.302  7.282  1 U 1.55e+008       0 e 0   37   37
109  7.115  7.283  1 U 2.09e+006       0 e 0   40   37
110  7.299  7.089  1 U 1.71e+006       0 e 0   37   40
```

```
111  7.301  7.197  1 U 1.34e+007       0 e 0   39   38
112  7.115  7.199  1 U  2.1e+007       0 e 0   40   38
113  8.344  8.285  1 U 4.01e+006       0 e 0   41   48
114  7.859  7.838  1 U 1.92e+008       0 e 0   53   53
115  8.191  7.838  1 U 8.88e+005       0 e 0   61   53
116  8.191  8.171  1 U 1.32e+008       0 e 0   61   61
117  7.859  8.173  1 U  6.7e+005       0 e 0   53   61
118  8.293  8.005  1 U -1.82e+005       0 e 0   72   75
119  8.139  8.118  1 U 1.35e+008       0 e 0   68   68
120  8.048  8.118  1 U 3.27e+006       0 e 0   72   68
121  8.138  7.789  1 U 3.38e+006       0 e 0   68   65
122  7.811  7.789  1 U 9.64e+007       0 e 0   65   65
123  7.809  8.118  1 U 3.58e+006       0 e 0   65   68
124  7.995  8.021  1 U -1.68e+006       0 e 0   75   75
125  4.504  7.284  1 U 2.74e+006       0 e 0   62   39
126  9.764  8.031  1 U 2.11e+005       0 e 0   36   75
127  4.305  8.030  1 U 3.06e+006       0 e 0   32   31
128  3.991  8.031  1 U  2.1e+006       0 e 0   26   31
129  3.198  8.030  1 U 4.87e+006       0 e 0   33   31
130  3.541  8.030  1 U 5.32e+006       0 e 0   34   31
131  0.922  8.073  1 U 2.08e+006       0 e 0   24   18
132  1.411  8.073  1 U 2.37e+006       0 e 0   23   18
133  1.642  8.073  1 U 2.21e+006       0 e 0   22   18
134  1.945  8.073  1 U  5.8e+006       0 e 0   20   18
135  3.091  8.073  1 U 2.61e+006       0 e 0   15   18
136  3.073  8.171  1 U 7.66e+005       0 e 0   64   61
137  3.506  8.171  1 U 1.91e+006       0 e 0   42   61
138  4.319  8.119  1 U 2.77e+006       0 e 0   69   68
139  4.300  8.171  1 U 2.08e+006       0 e 0   58   61
140  4.469  8.118  1 U 1.77e+006       0 e 0   66   68
141  4.503  7.789  1 U 1.73e+006       0 e 0   62   65
142  4.467  7.788  1 U 1.32e+006       0 e 0   66   65
143  3.927  7.789  1 U 1.73e+006       0 e 0   67   65
144  3.643  7.789  1 U 9.22e+005       0 e 0   63   65
145  3.922  8.119  1 U 1.92e+006       0 e 0   70   68
146  3.714  8.118  1 U 2.77e+006       0 e 0   71   68
147  8.339  3.486  1 U 1.91e+006       0 e 0   41   42
148  8.303  3.485  1 U 1.47e+006       0 e 0   48   42
```

```
149  8.192  3.689  1 U 3.99e+005       0 e 0   61   42
150  3.830  8.282  1 U 3.21e+005       0 e 0   49   48
151  3.998  8.283  1 U 3.96e+006       0 e 0   26   48
152  3.506  8.284  1 U 1.78e+006       0 e 0   42   48
153  0.859  3.485  1 U 3.31e+006       0 e 0   46   42
154  0.904  3.488  1 U    1e+006       0 e 0   45   42
155  8.340  3.520  1 U 2.14e+006       0 e 0   41   34
156  3.506  8.319  1 U 2.25e+006       0 e 0   42   41
157  3.540  8.319  1 U 2.57e+006       0 e 0   34   41
158  3.199  8.319  1 U 8.24e+005       0 e 0   33   41
159  8.340  3.170  1 U 8.04e+005       0 e 0   41   33
160  4.300  8.319  1 U 1.74e+006       0 e 0   32   41
161  3.830  8.319  1 U 1.19e+006       0 e 0   19   41
162  1.786  8.319  1 U 3.91e+006       0 e 0   43   41
163  1.550  8.319  1 U 2.82e+006       0 e 0   47   41
164  1.467  8.319  1 U 1.89e+006       0 e 0   44   41
165  0.908  8.319  1 U 7.94e+005       0 e 0   45   41
166  0.860  8.319  1 U 7.08e+005       0 e 0   46   41
167  7.113  0.840  1 U 1.97e+006       0 e 0   40   46
168  7.220  0.840  1 U  1.5e+006       0 e 0   37   46
169  6.818  6.798  1 U 3.18e+008       0 e 0   17   17
170  7.029  6.799  1 U 3.47e+007       0 e 0   16   17
171  6.819  7.011  1 U 3.62e+007       0 e 0   17   16
172  7.029  7.008  1 U 3.05e+008       0 e 0   16   16
173  0.905  6.798  1 U 2.82e+006       0 e 0   45   17
174  0.859  6.798  1 U 1.59e+006       0 e 0   46   17
175  0.913  7.009  1 U 1.81e+006       0 e 0   45   16
176  0.860  7.010  1 U 5.98e+005       0 e 0   46   16
177  7.114  6.802  1 U 1.55e+006       0 e 0   40   17
178  7.218  6.799  1 U 7.55e+005       0 e 0   37   17
179  6.819  7.205  1 U 1.19e+006       0 e 0   17   37
180  7.114  7.094  1 U 3.57e+008       0 e 0   40   40
181  6.820  7.101  1 U 1.95e+006       0 e 0   17   40
182  7.214  7.193  1 U  3.3e+008       0 e 0    5    5
183  3.043  3.158  1 U 1.51e+007       0 e 0    3    2
184  3.170  3.019  1 U 1.69e+007       0 e 0    2    3
185  7.828  3.166  1 U 1.05e+006       0 e 0    4    2
186  7.828  3.015  1 U 8.16e+005       0 e 0    4    3
```

```
187  4.356  3.158  1 U 1.93e+006        0 e 0    1    2
188  4.355  3.008  1 U 1.36e+006        0 e 0    1    3
189  7.115  3.173  1 U 1.64e+006        0 e 0   35   33
190  7.115  3.523  1 U  6.9e+005        0 e 0   35   34
191  7.221  3.520  1 U 2.27e+006        0 e 0   37   34
192  7.218  3.172  1 U    1e+006        0 e 0   37   33
193  3.921  3.688  1 U 1.64e+007        0 e 0   70   71
194  4.729  7.419  1 U 1.32e+006        0 e 0   54   57
195  8.304  3.809  1 U 3.36e+005        0 e 0   48   49
196  8.303  3.977  1 U  2.5e+006        0 e 0   48   26
197  8.051  3.969  1 U 1.36e+006        0 e 0   31   26
198  7.900  4.045  1 U      -106        0 e 0   53   49
199  7.856  3.973  1 U 2.41e+006        0 e 0   25   26
200  3.541  4.165  1 U 9.09e+005        0 e 0   34   14
201  3.196  4.165  1 U 4.24e+005        0 e 0   33   14
202  3.090  4.172  1 U 2.81e+006        0 e 0   15   14
203  4.388  8.258  1 U -2.46e+004        0 e 0   82   81
204  1.872  8.284  1 U 6.88e+006        0 e 0   52   48
205  1.786  8.284  1 U 2.14e+006        0 e 0   43   48
206  1.547  8.284  1 U 1.36e+006        0 e 0   47   48
207  1.462  8.284  1 U 1.88e+006        0 e 0   44   48
208  1.787  8.010  1 U 2.68e+006        0 e 0   76   75
209  1.694  8.010  1 U 1.57e+006        0 e 0   80   75
210  2.119  8.030  1 U 1.85e+006        0 e 0   29   31
211  2.126  7.832  1 U 4.22e+006        0 e 0   29   25
212  1.945  7.832  1 U 2.09e+006        0 e 0   20   25
213  2.404  7.832  1 U 1.42e+006        0 e 0   30   25
214  4.187  8.073  1 U 2.03e+006        0 e 0   14   18
215  4.729  7.840  1 U 1.98e+006        0 e 0   54   53
216  3.830  8.072  1 U 2.81e+006        0 e 0   19   18
217  4.187  7.009  1 U 2.73e+006        0 e 0   14   16
218  3.096  7.008  1 U 7.14e+006        0 e 0   15   16
219  3.872  8.027  1 U 3.94e+006        0 e 0   73   75
220  4.184  8.028  1 U 3.06e+006        0 e 0   74   75
221  4.298  7.839  1 U  7.8e+005        0 e 0   32   53
222  1.913  8.010  1 U 1.08e+006        0 e 0   77   75
223  0.920  7.832  1 U 1.62e+006        0 e 0   24   25
224  0.916  8.284  1 U  5.8e+005        0 e 0   45   48
```

```
225  8.470  7.004  1 U  8.1e+005       0 e 0   13   16
226  3.541  7.200  1 U 3.23e+006       0 e 0   34   37
227  1.550  7.200  1 U 2.11e+006       0 e 0   47   37
228  0.859  7.200  1 U 2.32e+006       0 e 0   46   37
229  0.914  7.196  1 U 1.21e+006       0 e 0   45   37
230  0.859  7.093  1 U  2.8e+006       0 e 0   46   40
231  0.908  7.094  1 U 1.14e+006       0 e 0   45   40
232  1.940  3.818  1 U 1.75e+006       0 e 0   20   19
233  0.920  3.809  1 U 4.69e+006       0 e 0   24   19
234  1.412  3.809  1 U 1.21e+006       0 e 0   23   19
235  1.639  3.809  1 U 6.56e+005       0 e 0   22   19
236  1.644  1.928  1 U 1.53e+006       0 e 0   22   20
237  1.417  1.924  1 U  8.7e+005       0 e 0   23   20
238  0.921  1.929  1 U 4.51e+006       0 e 0   24   20
239  3.830  1.918  1 U 1.41e+006       0 e 0   19   20
240  7.852  1.925  1 U 1.62e+006       0 e 0   25   20
241  8.094  1.925  1 U 4.58e+006       0 e 0   18   20
242  1.411  1.626  1 U 1.55e+007       0 e 0   23   22
243  1.642  1.387  1 U 1.51e+007       0 e 0   22   23
244  3.829  1.389  1 U 1.07e+006       0 e 0   19   23
245  3.829  1.618  1 U 4.33e+005       0 e 0   19   22
246  8.094  1.389  1 U 1.51e+006       0 e 0   18   23
247  8.094  1.624  1 U 1.55e+006       0 e 0   18   22
248  1.940  1.390  1 U 6.92e+005       0 e 0   20   23
249  1.942  1.626  1 U 1.66e+006       0 e 0   20   22
250  0.927  1.621  1 U 4.53e+006       0 e 0   21   22
251  0.936  1.378  1 U 1.14e+006       0 e 0   21   23
252  1.546  1.439  1 U 8.46e+006       0 e 0   47   44
253  1.458  1.526  1 U 1.08e+007       0 e 0   44   47
254  3.506  1.529  1 U 6.67e+005       0 e 0   42   47
255  3.505  1.443  1 U  1.9e+006       0 e 0   42   44
256  3.999  1.524  1 U 5.91e+005       0 e 0   49   50
257  3.999  1.426  1 U 8.02e+005       0 e 0   49   51
258  3.830  1.529  1 U 3.86e+005       0 e 0   19   47
259  3.829  1.442  1 U 6.32e+005       0 e 0   19   44
260  8.340  1.530  1 U 2.21e+006       0 e 0   41   47
261  8.340  1.442  1 U 1.44e+006       0 e 0   41   44
262  8.303  1.526  1 U  9.8e+005       0 e 0   41   47
```

```
263  8.304  1.439  1 U 1.28e+006       0 e 0   48   44
264  1.785  1.528  1 U 2.79e+006       0 e 0   43   47
265  1.786  1.440  1 U 1.25e+007       0 e 0   43   44
266  1.736  1.493  1 U 1.76e+007       0 e 0    9   10
267  2.115  2.394  1 U 4.67e+006       0 e 0   29   30
268  7.853  2.387  1 U 1.19e+006       0 e 0   25   30
269  3.990  2.388  1 U 1.15e+006       0 e 0   26   30
270  3.996  1.854  1 U 5.65e+006       0 e 0   49   52
271  8.304  1.852  1 U 5.26e+006       0 e 0   48   52
272  7.859  1.852  1 U 1.81e+006       0 e 0   53   52
273  2.401  2.100  1 U  3.5e+006       0 e 0   30   29
274  3.989  2.102  1 U 2.57e+006       0 e 0   26   29
275  7.852  2.104  1 U  3.1e+006       0 e 0   25   29
276  8.050  2.098  1 U 1.46e+006       0 e 0   31   29
277  4.267  1.716  1 U 1.77e+006       0 e 0    8    9
278  4.268  1.495  1 U 2.02e+006       0 e 0    8   10
279  3.508  1.767  1 U 2.09e+006       0 e 0   42   43
280  8.340  1.767  1 U 2.91e+006       0 e 0   41   43
281  8.303  1.767  1 U 1.56e+006       0 e 0   48   43
282  1.467  1.772  1 U 1.19e+007       0 e 0   44   43
283  1.548  1.774  1 U 2.06e+006       0 e 0   47   43
284  0.957  1.494  1 U 1.75e+006       0 e 0   11   10
285  0.905  1.544  1 U 1.64e+006       0 e 0   45   47
286  0.931  1.448  1 U 5.59e+006       0 e 0   45   44
287  0.861  1.526  1 U 3.67e+006       0 e 0   46   47
288  0.860  1.448  1 U 2.52e+006       0 e 0   46   44
289  0.859  1.770  1 U 9.32e+005       0 e 0   46   43
290  0.905  1.769  1 U 2.22e+006       0 e 0   45   43
291  0.957  1.716  1 U 4.85e+006       0 e 0   11    9
292  0.909  1.718  1 U 3.28e+006       0 e 0   12    9
293  0.907  1.521  1 U 3.63e+006       0 e 0   12   10
294  1.517  1.721  1 U 1.72e+007       0 e 0   10    9
295  8.907  1.717  1 U 4.56e+006       0 e 0    7    9
296  8.907  1.493  1 U 8.49e+005       0 e 0    7   10
297  8.473  1.717  1 U 1.12e+006       0 e 0   13    9
298  8.473  1.494  1 U 5.05e+005       0 e 0   13   10
299  8.046  3.858  1 U 2.68e+006       0 e 0   72   73
300  8.094  4.167  1 U 9.03e+005       0 e 0   72   74
```

```
301  8.026  3.857  1 U 1.08e+006        0 e 0   75   73
302  8.050  4.170  1 U 1.32e+006        0 e 0   75   74
303  7.853  3.809  1 U 1.96e+006        0 e 0   25   19
304  8.170  3.787  1 U 3.64e+005        0 e 0   81   84
305  8.339  4.276  1 U 6.34e+005        0 e 0   41   32
306  8.340  3.809  1 U 9.73e+005        0 e 0   41   19
307  8.093  3.810  1 U  2.2e+006        0 e 0   18   19
308  7.860  2.857  1 U 2.44e+006        0 e 0   53   55
309  7.859  3.045  1 U 1.32e+006        0 e 0   53   56
310  8.093  3.076  1 U  2.1e+006        0 e 0   18   15
311  8.473  3.079  1 U 5.46e+006        0 e 0   13   15
312  7.028  3.078  1 U  5.1e+006        0 e 0   16   15
313  4.187  3.059  1 U 3.49e+006        0 e 0   14   15
314  3.073  2.922  1 U 2.65e+006        0 e 0   56   55
315  2.953  3.051  1 U 1.69e+006        0 e 0   55   56
316  4.725  3.038  1 U -1.09e+006        0 e 0   54   56
317  4.727  2.851  1 U 2.72e+006        0 e 0   54   55
318  4.300  2.858  1 U 9.69e+005        0 e 0   32   55
319  1.737  4.251  1 U 1.53e+006        0 e 0    9    8
320  1.517  4.248  1 U 9.57e+005        0 e 0   10    8
321  0.907  4.246  1 U 1.24e+006        0 e 0   12    8
322  8.169  3.870  1 U 4.47e+005        0 e 0   81   83
323  8.140  3.904  1 U 1.28e+006        0 e 0   68   70
324  7.809  3.880  1 U 1.15e+006        0 e 0   65   67
325  8.138  3.691  1 U 1.86e+006        0 e 0   68   71
326  7.809  3.622  1 U 7.32e+005        0 e 0   65   63
327  4.317  3.901  1 U 3.18e+006        0 e 0   69   70
328  4.317  3.692  1 U 3.53e+006        0 e 0   69   71
329  3.714  3.905  1 U 1.32e+007        0 e 0   71   70
330  3.643  3.621  1 U 1.02e+008        0 e 0   63   63
331  3.833  1.764  1 U 1.28e+006        0 e 0   19   43
332  8.170  4.346  1 U 2.32e+006        0 e 0   81   82
333  8.142  4.292  1 U 9.31e+005        0 e 0   61   58
334  3.643  8.171  1 U    7e+005        0 e 0   63   61
335  3.925  8.256  1 U  4.8e+004        0 e 0   83   81
336  3.685  8.251  1 U -1.28e+004        0 e 0   84   81
337  3.823  4.388  1 U -1.93e+004        0 e 0   84   82
338  1.112  4.278  1 U  8.7e+005        0 e 0   60   59
```

```
339  8.140  3.190  1 U  3.1e+004       0 e 0   61   64
340  8.191  3.623  1 U 6.12e+005       0 e 0   61   63
341  8.900  3.167  1 U 8.69e+005       0 e 0    7    2
342  8.901  3.040  1 U 1.72e+005       0 e 0    7    3
343  7.323  3.247  1 U 9.05e+005       0 e 0   78   79
344  3.266  7.302  1 U  1.3e+006       0 e 0   79   78
345  1.466  1.530  1 U 8.07e+006       0 e 0   51   50
346  1.542  1.432  1 U 1.01e+007       0 e 0   50   51
347  4.299  1.089  1 U 2.22e+006       0 e 0   59   60
348  1.112  1.092  1 U 5.25e+008       0 e 0   60   60
349  0.927  1.621  1 U 4.53e+006       0 e 0   21   22
350  0.859  1.770  1 U 9.32e+005       0 e 0   46   43
351  0.905  1.769  1 U 2.22e+006       0 e 0   45   43
352  0.957  1.716  1 U 4.85e+006       0 e 0   11    9
353  0.909  1.718  1 U 3.28e+006       0 e 0   12    9
354  7.439  3.978  1 U 2.42e+005       0 e 0   57   49
355  7.439  3.484  1 U  2.7e+005       0 e 0   57   42
356  7.440  2.850  1 U 2.28e+005       0 e 0   57   55
357  7.439  3.044  1 U 1.83e+005       0 e 0   57   56
358  2.401  8.028  1 U 2.37e+005       0 e 0   30   31
359  1.112  8.171  1 U 7.04e+005       0 e 0   60   61
360  1.112  7.838  1 U 3.03e+005       0 e 0   60   65
361  1.113  7.788  1 U 2.25e+005       0 e 0   60   53
362  1.113  8.283  1 U 1.35e+005       0 e 0   60   48
363  0.860  8.284  1 U 3.67e+005       0 e 0   46   48
364  1.113  4.137  1 U 3.29e+006       0 e 0   60   58
365  1.114  3.980  1 U 2.26e+005       0 e 0   60   49
366  1.111  3.486  1 U 1.46e+005       0 e 0   60   42
367  1.112  3.622  1 U 1.51e+005       0 e 0   60   63
368  1.115  3.910  1 U 1.25e+005       0 e 0   60   70
369  1.113  3.043  1 U 1.37e+005       0 e 0   60   56
```

Trp-cage D-threonine shift list:

1	4.350	0.008	HA	1
2	3.165	0.005	HB2	1
3	3.031	0.015	HB3	1
4	7.818	0.010	HD21	1
5	7.202	0.011	HD22	1
6	3.174	0.000	QB	1
7	8.895	0.009	H	2
8	4.260	0.010	HA	2
9	1.724	0.009	HB2	2
10	1.506	0.012	HB3	2
11	0.957	0.000	HD1	2
12	0.909	0.002	HD2	2
13	8.462	0.010	H	3
14	4.181	0.013	HA	3
15	3.083	0.012	QB	3
16	7.016	0.010	QD	3
17	6.807	0.010	QE	3
18	8.081	0.010	H	4
19	3.821	0.010	HA	4
20	1.934	0.010	HB	4
21	0.930	0.004	HD1	4
22	1.630	0.010	HG12	4
23	1.398	0.014	HG13	4
24	0.916	0.010	HG2	4
25	7.842	0.010	H	5
26	3.985	0.010	HA	5
27	7.464	0.010	HE21	5
28	6.945	0.010	HE22	5
29	2.109	0.010	QB	5
30	2.398	0.007	QG	5
31	8.039	0.010	H	6
32	4.291	0.013	HA	6
33	3.185	0.013	HB2	6
34	3.532	0.010	HB3	6
35	7.110	0.010	HD1	6
36	9.723	0.036	HE1	6

```
37    7.231  0.038    HE3 6
38    7.207  0.010    HH2 6
39    7.295  0.009    HZ2 6
40    7.103  0.010    HZ3 6
41    8.328  0.011      H 7
42    3.511  0.052     HA 7
43    1.774  0.008    HB2 7
44    1.450  0.010    HB3 7
45    0.910  0.008    HD1 7
46    0.857  0.007    HD2 7
47    1.537  0.010     HG 7
48    8.292  0.010      H 8
49    3.941  0.085     HA 8
50    1.532  0.007    HG2 8
51    1.441  0.018    HG3 8
52    1.858  0.008     QB 8
53    7.848  0.021      H 9
54    4.728  0.002     HA 9
55    2.881  0.038    HB2 9
56    3.049  0.011    HB3 9
57    7.457  0.099      H 10
58    4.262  0.063     HA 10
59    4.289  0.011     HB 10
60    1.109  0.008    QG2 10
61    8.174  0.016      H 11
62    4.503  0.000     HA 12
63    3.631  0.010    HD2 12
64    3.098  0.054    HD3 12
65    7.801  0.015      H 13
66    4.468  0.001     HA 13
67    3.904  0.023     QB 13
68    8.128  0.010      H 14
69    4.304  0.011     HA 14
70    3.912  0.008    HB2 14
71    3.702  0.012    HB3 14
72    8.097  0.083      H 15
73    3.862  0.010    HA2 15
74    4.172  0.007    HA3 15
```

```
75    8.020  0.014      H 16
76    1.787  0.000    HB2 16
77    1.913  0.000    HB3 16
78    7.313  0.010     HE 16
79    3.257  0.009     QD 16
80    1.694  0.000     QG 16
81    8.212  0.043      H 20
82    4.374  0.020     HA 20
83    3.898  0.028    HB2 20
84    3.765  0.058    HB3 20
```

Trp-cage D-allo-threonine peak list:

```
# Number of dimensions 2
   1  7.565  1.079  1 U 2.45e+006        0 e 0   58   61
   2  7.900  7.563  1 U 3.26e+006        0 e 0   54   58
   3  8.140  7.563  1 U 7.61e+006        0 e 0   62   58
   4  4.294  7.567  1 U 2.54e+006        0 e 0   59   58
   5  1.082  7.563  1 U 2.75e+006        0 e 0   61   58
   6  7.566  4.288  1 U 1.25e+006        0 e 0   58   59
   7  8.163  4.290  1 U 1.82e+006        0 e 0   69   70
   8  8.095  4.294  1 U 1.08e+006        0 e 0   73   70
   9  3.915  4.294  1 U 2.99e+006        0 e 0   71   70
  10  3.683  4.293  1 U 2.93e+006        0 e 0   72   70
  11  8.266  7.899  1 U 4.52e+006        0 e 0   48   54
  12  7.564  7.899  1 U 3.11e+006        0 e 0   58   54
  13  2.947  7.899  1 U 3.87e+006        0 e 0   56   54
  14  3.067  7.899  1 U 1.55e+006        0 e 0   57   54
  15  7.564  8.139  1 U 7.12e+006        0 e 0   58   62
  16  4.295  8.095  1 U 1.98e+006        0 e 0   70   73
  17  4.172  8.096  1 U 2.44e+006        0 e 0   75   73
  18  3.877  8.095  1 U 4.78e+006        0 e 0   74   73
  19  7.563  7.561  1 U 1.45e+008        0 e 0   58   58
  20  8.165  8.094  1 U 2.99e+006        0 e 0   69   73
  21  8.096  8.094  1 U 1.78e+008        0 e 0   73   73
  22  8.031  8.094  1 U 4.25e+006        0 e 0   76   73
  23  4.176  3.872  1 U 5.95e+006        0 e 0   75   74
  24  3.873  4.186  1 U 1.09e+007        0 e 0   74   75
  25  8.901  8.457  1 U  3.2e+006        0 e 0    7   13
  26  8.459  8.900  1 U 2.87e+006        0 e 0   13    7
  27  8.901  8.900  1 U 1.53e+008        0 e 0    7    7
  28  4.359  8.900  1 U 8.56e+006        0 e 0    1    7
  29  4.268  8.902  1 U 2.35e+006        0 e 0    8    7
  30  4.268  8.458  1 U 3.07e+006        0 e 0    8   13
  31  4.214  8.456  1 U 2.81e+006        0 e 0   14   13
  32  8.069  8.458  1 U  4.5e+006        0 e 0   18   13
  33  7.036  8.458  1 U 6.75e+005        0 e 0   16   13
  34  3.163  8.900  1 U 1.98e+006        0 e 0    6    7
  35  3.092  8.458  1 U 9.89e+006        0 e 0   15   13
```

```
36  1.729  8.900  1 U 7.61e+006        0 e 0    9    7
37  1.515  8.900  1 U 1.18e+006        0 e 0   10    7
38  1.728  8.458  1 U 1.72e+006        0 e 0    9   13
39  1.513  8.458  1 U 8.51e+005        0 e 0   10   13
40  0.958  8.900  1 U 6.16e+005        0 e 0   11    7
41  0.910  8.900  1 U  7.7e+005        0 e 0   12    7
42  0.911  8.458  1 U 6.09e+005        0 e 0   12   13
43  1.938  0.908  1 U 6.25e+006        0 e 0   20   24
44  8.071  8.069  1 U 1.86e+008        0 e 0   18   18
45  8.459  8.069  1 U 4.13e+006        0 e 0   13   18
46  7.863  8.069  1 U 3.57e+006        0 e 0   25   18
47  8.048  7.863  1 U 4.74e+006        0 e 0   31   25
48  8.068  7.863  1 U 4.14e+006        0 e 0   18   25
49  7.863  7.863  1 U 1.56e+008        0 e 0   25   25
50  7.864  8.045  1 U 4.53e+006        0 e 0   25   31
51  8.459  8.458  1 U 1.75e+008        0 e 0   13   13
52  8.308  8.046  1 U 4.01e+006        0 e 0   41   31
53  4.047  7.899  1 U 2.28e+006        0 e 0   49   54
54  4.268  7.862  1 U 8.86e+005        0 e 0    8   25
55  4.012  7.863  1 U 3.13e+006        0 e 0   26   25
56  3.836  7.863  1 U 3.39e+006        0 e 0   19   25
57  3.722  3.191  1 U 1.88e+007        0 e 0   64   65
58  4.325  3.205  1 U 3.56e+006        0 e 0   32   33
59  8.046  3.199  1 U 4.43e+006        0 e 0   31   33
60  8.046  3.528  1 U    5e+006        0 e 0   31   34
61  4.322  3.532  1 U 1.83e+006        0 e 0   32   34
62  3.525  3.202  1 U  1.5e+007        0 e 0   34   33
63  3.206  3.530  1 U 1.61e+007        0 e 0   33   34
64  3.193  3.722  1 U 1.71e+007        0 e 0   65   64
65  4.213  3.527  1 U 2.01e+006        0 e 0   14   34
66  4.213  3.196  1 U 6.48e+005        0 e 0   14   33
67  3.530  4.319  1 U 2.01e+006        0 e 0   34   32
68  3.205  4.323  1 U 2.26e+006        0 e 0   33   32
69  8.046  4.322  1 U 1.39e+006        0 e 0   31   32
70  8.045  8.047  1 U  1.8e+008        0 e 0   31   31
71  8.046  8.307  1 U 4.39e+006        0 e 0   31   41
72  7.244  8.307  1 U 2.02e+006        0 e 0   37   41
73  8.267  8.306  1 U 4.74e+006        0 e 0   48   41
```

```
 74  8.262  8.262  1 U 3.33e+008       0 e 0   48   48
 75  7.900  8.265  1 U 4.06e+006       0 e 0   54   48
 76  8.308  8.308  1 U 1.73e+008       0 e 0   41   41
 77  3.167  4.365  1 U 1.34e+006       0 e 0    2    1
 78  3.049  4.368  1 U  6.5e+005       0 e 0    3    1
 79  8.901  4.357  1 U 3.14e+006       0 e 0    7    1
 80  8.901  4.263  1 U 9.49e+005       0 e 0    7    8
 81  8.459  4.266  1 U 1.24e+006       0 e 0   13    8
 82  8.458  4.214  1 U 1.58e+006       0 e 0   13   14
 83  8.069  4.212  1 U 1.36e+006       0 e 0   18   14
 84  7.037  4.212  1 U 1.19e+006       0 e 0   16   14
 85  9.763  9.762  1 U 3.69e+007       0 e 0   36   36
 86  7.118  9.762  1 U 2.34e+006       0 e 0   35   36
 87  7.304  9.763  1 U 8.93e+005       0 e 0   39   36
 88  9.764  7.303  1 U 3.13e+006       0 e 0   36   39
 89  9.764  7.118  1 U 9.24e+006       0 e 0   36   35
 90  8.308  7.244  1 U  2.4e+006       0 e 0   41   37
 91  4.328  7.118  1 U 2.37e+006       0 e 0   32   35
 92  3.205  7.118  1 U    4e+006       0 e 0   33   35
 93  3.526  7.117  1 U 1.33e+006       0 e 0   34   35
 94  7.830  7.829  1 U 1.48e+008       0 e 0    4    4
 95  7.829  7.211  1 U 5.74e+007       0 e 0    4    5
 96  7.211  7.829  1 U 5.75e+007       0 e 0    5    4
 97  7.221  7.223  1 U 2.32e+008       0 e 0   38   38
 98  7.450  7.449  1 U 9.21e+007       0 e 0   27   27
 99  7.450  6.960  1 U 3.98e+007       0 e 0   27   28
100  6.960  7.449  1 U 4.08e+007       0 e 0   28   27
101  6.961  6.961  1 U 8.35e+007       0 e 0   28   28
102  3.162  7.829  1 U 1.83e+006       0 e 0    2    4
103  3.046  7.829  1 U 1.19e+006       0 e 0    3    4
104  2.413  7.449  1 U 6.17e+005       0 e 0   30   27
105  7.304  7.303  1 U 1.65e+008       0 e 0   39   39
106  7.120  7.119  1 U 3.79e+008       0 e 0   35   35
107  7.226  7.302  1 U 1.87e+007       0 e 0   38   39
108  7.222  7.125  1 U  1.1e+007       0 e 0   38   40
109  7.240  7.245  1 U 1.48e+008       0 e 0   37   37
110  7.130  7.245  1 U 1.57e+007       0 e 0   40   37
111  7.247  7.124  1 U 1.53e+007       0 e 0   37   40
```

```
112  7.301  7.224  1 U 1.49e+007       0 e 0   39   38
113  7.130  7.223  1 U 1.04e+007       0 e 0   40   38
114  8.312  8.263  1 U 2.19e+006       0 e 0   41   48
115  7.900  7.899  1 U 1.39e+008       0 e 0   54   54
116  8.140  7.897  1 U 8.06e+005       0 e 0   62   54
117  8.139  8.139  1 U 1.68e+008       0 e 0   62   62
118  7.900  8.137  1 U 7.83e+005       0 e 0   54   62
119  8.097  8.028  1 U 3.22e+006       0 e 0   73   76
120  8.163  8.162  1 U 1.69e+008       0 e 0   69   69
121  8.094  8.162  1 U 3.83e+006       0 e 0   73   69
122  8.163  7.932  1 U  4.6e+006       0 e 0   69   66
123  7.933  7.933  1 U  1.4e+008       0 e 0   66   66
124  7.933  8.162  1 U 4.56e+006       0 e 0   66   69
125  8.034  8.035  1 U 1.96e+008       0 e 0   76   76
126  4.516  7.303  1 U 2.99e+006       0 e 0   63   39
127  9.764  8.031  1 U 6.14e+005       0 e 0   36   76
128  4.326  8.045  1 U 3.25e+006       0 e 0   32   31
129  4.012  8.046  1 U 2.65e+006       0 e 0   26   31
130  3.205  8.045  1 U 5.62e+006       0 e 0   33   31
131  3.528  8.045  1 U 5.92e+006       0 e 0   34   31
132  0.915  8.069  1 U 2.31e+006       0 e 0   24   18
133  1.400  8.069  1 U 2.55e+006       0 e 0   23   18
134  1.632  8.069  1 U 2.31e+006       0 e 0   22   18
135  1.939  8.069  1 U 6.24e+006       0 e 0   20   18
136  3.090  8.069  1 U 3.62e+006       0 e 0   15   18
137  3.193  8.139  1 U 7.45e+005       0 e 0   65   62
138  3.564  8.139  1 U 2.15e+006       0 e 0   42   62
139  4.296  8.165  1 U 3.69e+006       0 e 0   70   69
140  4.296  8.138  1 U 2.32e+006       0 e 0   59   62
141  4.484  8.163  1 U 2.34e+006       0 e 0   67   69
142  4.515  7.932  1 U 2.81e+006       0 e 0   63   66
143  4.486  7.931  1 U 2.12e+006       0 e 0   67   66
144  7.036  8.069  1 U 7.62e+005       0 e 0   16   18
145  3.925  7.932  1 U 5.04e+006       0 e 0   68   66
146  3.723  7.932  1 U 1.25e+006       0 e 0   64   66
147  3.921  8.161  1 U 2.52e+006       0 e 0   71   69
148  3.682  8.163  1 U 3.27e+006       0 e 0   72   69
149  8.309  3.562  1 U 2.47e+006       0 e 0   41   42
```

```
150  8.266  3.563  1 U 2.11e+006        0 e 0   48   42
151  8.140  3.563  1 U 1.93e+006        0 e 0   62   42
152  4.047  8.265  1 U 3.21e+006        0 e 0   49   48
153  4.011  8.265  1 U  1.1e+006        0 e 0   26   48
154  3.564  8.265  1 U 2.24e+006        0 e 0   42   48
155  0.854  3.563  1 U 3.99e+006        0 e 0   46   42
156  0.903  3.562  1 U 1.25e+006        0 e 0   45   42
157  8.308  3.528  1 U 2.54e+006        0 e 0   41   34
158  3.561  8.308  1 U 3.06e+006        0 e 0   42   41
159  3.529  8.307  1 U 3.22e+006        0 e 0   34   41
160  3.205  8.307  1 U 1.16e+006        0 e 0   33   41
161  8.308  3.195  1 U 9.57e+005        0 e 0   41   33
162  4.325  8.307  1 U  2.6e+006        0 e 0   32   41
163  3.836  8.307  1 U 1.47e+006        0 e 0   19   41
164  1.769  8.307  1 U 5.05e+006        0 e 0   43   41
165  1.551  8.307  1 U 3.57e+006        0 e 0   47   41
166  1.473  8.307  1 U 2.41e+006        0 e 0   44   41
167  0.906  8.307  1 U 1.19e+006        0 e 0   45   41
168  0.855  8.308  1 U 9.22e+005        0 e 0   46   41
169  7.125  0.854  1 U 2.22e+006        0 e 0   40   46
170  7.245  0.857  1 U 1.62e+006        0 e 0   37   46
171  6.823  6.823  1 U 3.65e+008        0 e 0   17   17
172  7.034  6.819  1 U 2.19e+007        0 e 0   16   17
173  6.826  7.040  1 U 2.07e+007        0 e 0   17   16
174  7.036  7.036  1 U 3.31e+008        0 e 0   16   16
175  0.903  6.823  1 U 2.85e+006        0 e 0   45   17
176  0.854  6.823  1 U 1.56e+006        0 e 0   46   17
177  0.909  7.036  1 U 2.03e+006        0 e 0   45   16
178  0.853  7.038  1 U 6.54e+005        0 e 0   46   16
179  7.126  6.823  1 U 1.23e+006        0 e 0   40   17
180  7.242  6.820  1 U  7.5e+005        0 e 0   37   17
181  6.823  7.243  1 U 9.94e+005        0 e 0   17   37
182  7.124  7.125  1 U 2.52e+008        0 e 0   40   40
183  6.821  7.124  1 U 2.09e+006        0 e 0   17   40
184  7.217  7.215  1 U 2.47e+008        0 e 0    5    5
185  3.045  3.166  1 U 1.86e+007        0 e 0    3    2
186  3.159  3.039  1 U 1.94e+007        0 e 0    2    3
187  7.830  3.166  1 U 1.07e+006        0 e 0    4    2
```

```
188  7.829  3.041  1 U 9.24e+005       0 e 0    4    3
189  4.358  3.165  1 U 1.07e+006       0 e 0    1    2
190  4.358  3.034  1 U 1.22e+006       0 e 0    1    3
191  7.119  3.200  1 U 2.21e+006       0 e 0   35   33
192  7.119  3.530  1 U 8.43e+005       0 e 0   35   34
193  7.245  3.527  1 U 2.66e+006       0 e 0   37   34
194  7.244  3.197  1 U 1.16e+006       0 e 0   37   33
195  3.914  3.676  1 U 2.12e+007       0 e 0   71   72
196  4.767  7.564  1 U 1.54e+006       0 e 0   55   58
197  8.265  4.046  1 U 2.48e+006       0 e 0   48   49
198  8.265  4.010  1 U 1.11e+006       0 e 0   48   26
199  8.047  4.010  1 U 1.95e+006       0 e 0   31   26
200  7.900  4.045  1 U 1.48e+006       0 e 0   54   49
201  7.864  4.010  1 U 1.98e+006       0 e 0   25   26
202  3.529  4.212  1 U 1.24e+006       0 e 0   34   14
203  3.204  4.210  1 U 6.26e+005       0 e 0   33   14
204  3.094  4.217  1 U 4.57e+006       0 e 0   15   14
205  4.388  8.258  1 U  7.8e+006       0 e 0   83   82
206  1.878  8.265  1 U 6.14e+006       0 e 0   53   48
207  1.769  8.265  1 U 2.28e+006       0 e 0   43   48
208  1.548  8.265  1 U 1.57e+006       0 e 0   47   48
209  1.469  8.265  1 U 2.05e+006       0 e 0   44   48
210  1.783  8.031  1 U 3.04e+006       0 e 0   77   76
211  1.691  8.030  1 U 1.82e+006       0 e 0   81   76
212  2.120  8.045  1 U 2.52e+006       0 e 0   29   31
213  2.123  7.863  1 U  5.8e+006       0 e 0   29   25
214  1.939  7.863  1 U 2.36e+006       0 e 0   20   25
215  2.407  7.862  1 U 1.55e+006       0 e 0   30   25
216  4.213  8.069  1 U 2.51e+006       0 e 0   14   18
217  4.766  7.898  1 U 2.14e+006       0 e 0   55   54
218  3.837  8.068  1 U 3.01e+006       0 e 0   19   18
219  4.213  7.036  1 U 2.96e+006       0 e 0   14   16
220  3.093  7.036  1 U 9.91e+006       0 e 0   15   16
221  3.876  8.030  1 U 1.41e+006       0 e 0   74   76
222  4.171  8.030  1 U 1.15e+006       0 e 0   75   76
223  4.326  7.898  1 U  7.8e+005       0 e 0   32   54
224  1.913  8.031  1 U 1.25e+006       0 e 0   78   76
225  0.913  7.862  1 U 1.98e+006       0 e 0   24   25
```

```
226  0.909  8.264  1 U 9.13e+005        0 e 0   45   48
227  8.458  7.039  1 U 8.95e+005        0 e 0   13   16
228  3.529  7.244  1 U 3.84e+006        0 e 0   34   37
229  1.551  7.244  1 U 2.39e+006        0 e 0   47   37
230  0.854  7.244  1 U 2.16e+006        0 e 0   46   37
231  0.906  7.242  1 U 9.48e+005        0 e 0   45   37
232  0.854  7.125  1 U 3.06e+006        0 e 0   46   40
233  0.904  7.122  1 U 1.11e+006        0 e 0   45   40
234  1.937  3.842  1 U 2.28e+006        0 e 0   20   19
235  0.913  3.834  1 U 6.23e+006        0 e 0   24   19
236  1.399  3.834  1 U 1.61e+006        0 e 0   23   19
237  1.634  3.836  1 U 7.49e+005        0 e 0   22   19
238  1.633  1.941  1 U 1.97e+006        0 e 0   22   20
239  1.404  1.937  1 U 9.08e+005        0 e 0   23   20
240  0.915  1.943  1 U 6.31e+006        0 e 0   24   20
241  3.836  1.932  1 U 2.02e+006        0 e 0   19   20
242  7.864  1.939  1 U 1.86e+006        0 e 0   25   20
243  8.070  1.939  1 U 5.39e+006        0 e 0   18   20
244  1.401  1.635  1 U 1.45e+007        0 e 0   23   22
245  1.632  1.395  1 U 1.38e+007        0 e 0   22   23
246  3.836  1.399  1 U 1.32e+006        0 e 0   19   23
247  3.836  1.631  1 U 7.05e+005        0 e 0   19   22
248  8.070  1.398  1 U 1.88e+006        0 e 0   18   23
249  8.070  1.634  1 U 1.85e+006        0 e 0   18   22
250  1.935  1.397  1 U 8.44e+005        0 e 0   20   23
251  1.938  1.629  1 U 1.71e+006        0 e 0   20   22
252  0.919  1.629  1 U 3.74e+006        0 e 0   21   22
253  0.919  1.397  1 U 1.32e+006        0 e 0   21   23
254  1.543  1.461  1 U 1.18e+007        0 e 0   47   44
255  1.469  1.547  1 U 1.36e+007        0 e 0   44   47
256  3.563  1.550  1 U    1e+006        0 e 0   42   47
257  3.564  1.473  1 U 2.04e+006        0 e 0   42   44
258  4.046  1.546  1 U 4.85e+005        0 e 0   49   51
259  4.047  1.453  1 U  7.7e+005        0 e 0   49   52
260  3.836  1.553  1 U 5.72e+005        0 e 0   19   47
261  3.836  1.468  1 U 8.37e+005        0 e 0   19   44
262  8.308  1.550  1 U 2.68e+006        0 e 0   41   47
263  8.308  1.467  1 U 1.73e+006        0 e 0   41   44
```

```
264  8.266  1.547  1 U  1.2e+006       0 e 0   41   47
265  8.266  1.464  1 U 1.55e+006       0 e 0   48   44
266  1.768  1.548  1 U 3.09e+006       0 e 0   43   47
267  1.769  1.466  1 U 1.35e+007       0 e 0   43   44
268  1.728  1.508  1 U 1.99e+007       0 e 0    9   10
269  2.116  2.416  1 U 4.97e+006       0 e 0   29   30
270  7.864  2.408  1 U 1.18e+006       0 e 0   25   30
271  4.011  2.406  1 U 1.11e+006       0 e 0   26   30
272  1.875  1.874  1 U 3.21e+008       0 e 0   53   53
273  3.285  1.874  1 U 4.15e+006       0 e 0   80   53
274  4.047  1.877  1 U 3.17e+006       0 e 0   49   53
275  8.265  1.878  1 U 4.98e+006       0 e 0   48   53
276  7.900  1.879  1 U 1.49e+006       0 e 0   54   53
277  2.398  2.116  1 U  4.5e+006       0 e 0   30   29
278  4.011  2.120  1 U 3.86e+006       0 e 0   26   29
279  7.864  2.121  1 U 4.33e+006       0 e 0   25   29
280  8.046  2.120  1 U 2.06e+006       0 e 0   31   29
281  4.268  1.729  1 U 2.49e+006       0 e 0    8    9
282  4.268  1.511  1 U 2.08e+006       0 e 0    8   10
283  3.562  1.766  1 U 1.57e+006       0 e 0   42   43
284  8.308  1.769  1 U 3.47e+006       0 e 0   41   43
285  8.266  1.769  1 U 1.86e+006       0 e 0   48   43
286  1.472  1.774  1 U 1.48e+007       0 e 0   44   43
287  1.548  1.773  1 U 2.75e+006       0 e 0   47   43
288  0.958  1.511  1 U  2.4e+006       0 e 0   11   10
289  0.905  1.544  1 U 4.89e+006       0 e 0   45   47
290  0.905  1.470  1 U 3.44e+006       0 e 0   45   44
291  0.856  1.547  1 U 4.59e+006       0 e 0   46   47
292  0.855  1.471  1 U 2.88e+006       0 e 0   46   44
293  0.854  1.772  1 U 1.16e+006       0 e 0   46   43
294  0.903  1.769  1 U 3.05e+006       0 e 0   45   43
295  0.958  1.729  1 U 5.98e+006       0 e 0   11    9
296  0.910  1.732  1 U 4.86e+006       0 e 0   12    9
297  0.907  1.511  1 U 2.82e+006       0 e 0   12   10
298  1.514  1.732  1 U 1.93e+007       0 e 0   10    9
299  1.729  1.729  1 U 2.33e+008       0 e 0    9    9
300  8.901  1.729  1 U 6.11e+006       0 e 0    7    9
301  8.900  1.509  1 U 1.15e+006       0 e 0    7   10
```

```
302  8.459  1.728  1 U 1.37e+006       0 e 0   13    9
303  8.458  1.509  1 U 6.39e+005       0 e 0   13   10
304  8.096  3.873  1 U 2.29e+006       0 e 0   73   74
305  8.095  4.173  1 U  5.6e+005       0 e 0   73   75
306  8.033  3.872  1 U 7.69e+005       0 e 0   76   74
307  8.032  4.174  1 U 5.32e+005       0 e 0   76   75
308  7.864  3.835  1 U 2.64e+006       0 e 0   25   19
309  8.262  3.831  1 U 6.92e+005       0 e 0   82   85
310  8.308  4.320  1 U    1e+006       0 e 0   41   32
311  8.308  3.835  1 U 1.23e+006       0 e 0   41   19
312  8.069  3.835  1 U 2.73e+006       0 e 0   18   19
313  7.900  2.941  1 U 2.48e+006       0 e 0   54   56
314  7.900  3.076  1 U 1.23e+006       0 e 0   54   57
315  8.069  3.091  1 U 3.36e+006       0 e 0   18   15
316  8.459  3.092  1 U 8.34e+006       0 e 0   13   15
317  7.037  3.092  1 U 7.45e+006       0 e 0   16   15
318  4.210  3.090  1 U 5.18e+006       0 e 0   14   15
319  3.069  2.941  1 U 2.02e+007       0 e 0   57   56
320  2.950  3.076  1 U  2.1e+007       0 e 0   56   57
321  4.768  3.076  1 U 6.31e+005       0 e 0   55   57
322  4.764  2.943  1 U 1.43e+006       0 e 0   55   56
323  4.323  2.939  1 U 9.44e+005       0 e 0   32   56
324  1.729  4.270  1 U 2.29e+006       0 e 0    9    8
325  1.514  4.270  1 U 1.27e+006       0 e 0   10    8
326  0.909  4.266  1 U 1.99e+006       0 e 0   12    8
327  8.258  3.920  1 U 6.99e+005       0 e 0   82   84
328  8.162  3.921  1 U 2.03e+006       0 e 0   69   71
329  7.933  3.924  1 U 3.67e+006       0 e 0   66   68
330  8.163  3.680  1 U 2.46e+006       0 e 0   69   72
331  7.933  3.721  1 U 8.92e+005       0 e 0   66   64
332  4.296  3.917  1 U 5.51e+006       0 e 0   70   71
333  4.295  3.679  1 U 5.29e+006       0 e 0   70   72
334  3.684  3.918  1 U 1.64e+007       0 e 0   72   71
335  3.724  3.722  1 U 9.07e+007       0 e 0   64   64
336  3.838  1.772  1 U 1.68e+006       0 e 0   19   43
337  8.259  4.383  1 U 2.94e+006       0 e 0   82   83
338  8.142  4.292  1 U    1e+006       0 e 0   62   59
339  3.723  8.139  1 U 5.79e+005       0 e 0   64   62
```

```
340  3.925  8.256  1 U 6.47e+005       0 e 0   84   82
341  3.829  8.261  1 U 7.31e+005       0 e 0   85   82
342  3.823  4.388  1 U 5.57e+005       0 e 0   85   83
343  1.082  4.308  1 U  7.1e+006       0 e 0   61   60
344  8.140  3.190  1 U 7.05e+005       0 e 0   62   65
345  8.140  3.722  1 U 6.05e+005       0 e 0   62   64
346  3.277  3.276  1 U 9.35e+007       0 e 0   80   80
347  8.900  3.167  1 U 1.15e+006       0 e 0    7    2
348  8.901  3.040  1 U 6.19e+005       0 e 0    7    3
349  3.286  3.206  1 U 4.13e+006       0 e 0   80   33
350  1.873  3.284  1 U 4.52e+006       0 e 0   53   80
351  7.323  3.262  1 U 1.26e+006       0 e 0   79   80
352  3.265  7.322  1 U 2.09e+006       0 e 0   80   79
353  1.452  1.545  1 U 9.93e+006       0 e 0   52   51
354  1.536  1.456  1 U 1.03e+007       0 e 0   51   52
355  4.302  1.080  1 U 1.13e+007       0 e 0   60   61
356  1.687  1.688  1 U 1.13e+008       0 e 0   81   81
357  3.564  7.563  1 U 4.12e+005       0 e 0   42   58
358  3.070  7.562  1 U 2.31e+005       0 e 0   57   58
359  1.883  7.563  1 U  2.8e+005       0 e 0   50   58
360  1.469  7.560  1 U  1.3e+005       0 e 0   44   58
361  1.082  7.898  1 U 3.12e+005       0 e 0   61   54
362  1.081  8.139  1 U 5.44e+005       0 e 0   61   69
363  1.084  8.166  1 U 2.23e+005       0 e 0   61   62
364  1.082  3.920  1 U 4.41e+005       0 e 0   61   68
365  1.082  3.680  1 U 1.73e+005       0 e 0   61   72
366  1.078  2.933  1 U 1.39e+005       0 e 0   61   56
367  1.082  1.081  1 U 6.15e+008       0 e 0   61   61
368  7.563  4.046  1 U 4.16e+005       0 e 0   58   49
369  7.563  3.562  1 U 3.62e+005       0 e 0   58   42
370  1.082  4.289  1 U 1.77e+006       0 e 0   61   59
371  1.081  3.067  1 U 1.75e+005       0 e 0   61   57
372  1.082  3.899  1 U 9.51e+004       0 e 0   61   71
373  1.082  7.931  1 U 1.88e+005       0 e 0   61   66
374  1.084  8.098  1 U 1.28e+005       0 e 0   61   73
375  4.268  2.121  1 U 1.44e+006       0 e 0    8   29
376  4.324  3.069  1 U 3.94e+005       0 e 0   32   57
377  4.326  7.244  1 U 5.27e+005       0 e 0   32   37
```

```
378  4.299  7.932  1 U 5.91e+005       0 e 0   70   66
379  4.360  8.460  1 U 3.08e+005       0 e 0    1   13
380  7.034  3.837  1 U 4.66e+005       0 e 0   16   19
381  7.036  3.528  1 U    2e+005       0 e 0   16   34
382  7.036  1.630  1 U  2.2e+005       0 e 0   16   22
383  7.037  1.504  1 U 4.51e+005       0 e 0   16   10
384  7.037  1.400  1 U 2.47e+005       0 e 0   16   23
385  7.564  2.928  1 U 2.28e+005       0 e 0   58   56
386  7.564  3.066  1 U 2.29e+005       0 e 0   58   57
387  7.568  8.270  1 U 2.78e+005       0 e 0   58   48
388  3.217  9.762  1 U 1.78e+005       0 e 0   33   36
389  1.783  9.762  1 U 2.57e+005       0 e 0   77   36
390  1.914  9.762  1 U 1.44e+005       0 e 0   78   36
391  1.697  9.763  1 U 1.23e+005       0 e 0   81   36
392  8.033  9.764  1 U 1.51e+005       0 e 0   76   36
393  3.807  4.242  1 U 1.47e+005       0 e 0   85    8
394  3.836  7.038  1 U 4.93e+005       0 e 0   19   16
395  4.385  3.111  1 U 1.85e+005       0 e 0   83   15
396  8.043  7.118  1 U 7.27e+005       0 e 0   31   35
397  8.046  7.250  1 U  6.8e+005       0 e 0   31   37
398  8.068  8.900  1 U 1.97e+005       0 e 0   18    7
399  3.281  7.118  1 U 9.03e+005       0 e 0   80   35
400  3.284  3.529  1 U 4.59e+005       0 e 0   80   34
```

Trp-cage D-allo-threonine shift list:

1	4.361	0.004	HA	1
2	3.165	0.003	HB2	1
3	3.042	0.005	HB3	1
4	7.829	0.000	HD21	1
5	7.213	0.003	HD22	1
6	3.163	0.000	QB	1
7	8.900	0.001	H	2
8	4.265	0.007	HA	2
9	1.729	0.001	HB2	2
10	1.511	0.003	HB3	2
11	0.958	0.000	HD1	2
12	0.909	0.001	HD2	2
13	8.458	0.001	H	3
14	4.213	0.002	HA	3
15	3.094	0.006	QB	3
16	7.037	0.001	QD	3
17	6.822	0.002	QE	3
18	8.069	0.001	H	4
19	3.836	0.002	HA	4
20	1.938	0.003	HB	4
21	0.919	0.000	HD1	4
22	1.632	0.002	HG12	4
23	1.399	0.002	HG13	4
24	0.913	0.003	HG2	4
25	7.863	0.001	H	5
26	4.011	0.001	HA	5
27	7.449	0.000	HE21	5
28	6.960	0.001	HE22	5
29	2.120	0.002	QB	5
30	2.408	0.006	QG	5
31	8.046	0.001	H	6
32	4.324	0.002	HA	6
33	3.203	0.005	HB2	6
34	3.528	0.002	HB3	6
35	7.118	0.001	HD1	6
36	9.763	0.001	HE1	6

```
37    7.244  0.002    HE3 6
38    7.223  0.002    HH2 6
39    7.303  0.001    HZ2 6
40    7.125  0.002    HZ3 6
41    8.306  0.008      H 7
42    3.563  0.001     HA 7
43    1.770  0.002    HB2 7
44    1.469  0.003    HB3 7
45    0.905  0.002    HD1 7
46    0.855  0.001    HD2 7
47    1.548  0.003     HG 7
48    8.265  0.002      H 8
49    4.046  0.001     HA 8
50    1.883  0.000    HB2 8
51    1.542  0.004    HG2 8
52    1.454  0.002    HG3 8
53    1.876  0.002     QB 8
54    7.899  0.001      H 9
55    4.766  0.001     HA 9
56    2.940  0.007    HB2 9
57    3.071  0.004    HB3 9
58    7.564  0.002      H 10
59    4.292  0.003     HA 10
60    4.305  0.003     HB 10
61    1.082  0.001    QG2 10
62    8.141  0.007      H 11
63    4.515  0.001     HA 12
64    3.722  0.001    HD2 12
65    3.192  0.001    HD3 12
66    7.932  0.001      H 13
67    4.485  0.001     HA 13
68    3.923  0.002     QB 13
69    8.161  0.006      H 14
70    4.295  0.002     HA 14
71    3.915  0.007    HB2 14
72    3.681  0.002    HB3 14
73    8.095  0.001      H 15
74    3.874  0.002    HA2 15
```

```
75    4.175  0.005    HA3 15
76    8.031  0.002      H 16
77    1.783  0.000    HB2 16
78    1.914  0.001    HB3 16
79    7.322  0.001     HE 16
80    3.278  0.008     QD 16
81    1.691  0.004     QG 16
82    8.259  0.002      H 20
83    4.386  0.002     HA 20
84    3.922  0.003    HB2 20
85    3.823  0.009    HB3 20
```

ProtEvolver C++ code:

```
//*************************************************************************************************************

//*************************************************************************************************************

//**********************************                          ****************************************

//**********************************     protEvolver 3.0      ****************************************

//**********************************                          ****************************************

//*************************************************************************************************************

//*******   -Folding Selective Protein Genetic Algorithm Based Evolution in Implicit Solvent -   ********

//*************************************************************************************************************


/////// Just specify infile structure, active chains and residues indexes, and it will evolve a sequence
favorable for folding


//--Included files and functions-----------------------------------------------------------------------

#include <iostream>

#include <string>

#include <time.h>

#include <dirent.h>

#include <sstream>

#include <fstream>

#include <unistd.h>

#include "ensemble.h"

#include "PDBInterface.h"


void randomizeSideChains(protein* _prot, UInt _chainIndex);

vector <UInt> getChainSequence(protein* _prot, UInt _chainIndex);

vector <UInt> getMutationPosition(protein* _prot, UInt* _activeChains, UInt _activeChainsSize, UInt*
_activeResidues, UInt _activeResiduesSize);

UInt getProbabilisticMutation(vector < vector < UInt > > &_sequencePool, vector <UInt> _mutantPosition, UInt
*_aminoacids, UInt aaSize);

vector < vector < UInt > > buildSequencePool();

int runmode;

int tInitialMut, tPostInitialMut, tInitialOpt, tPostInitialOpt, tMutStart, tMutFinished, tMutOptStart,
tMutOptFinished, randStart = 1;

int getCurrentTime();


//--Program setup--------------------------------------------------------------------------------------

int main (int argc, char* argv[]) {

         //--Running parameters

         if (argc < 4) {
```

```
                cout << "protEvolver <runmode> <inFile.pdb> <localoutfile.out>" << endl;
                cout << "runmode =  1 | Optimize system energy with random starting sequence" << endl;
                cout << "runmode =  2 | Optimize binding energy with random starting sequence" << endl;
                cout << "runmode =  3 | Optimize both system and binding energies with random starting
sequence" << endl;
                cout << "runmode = 11 | Optimize system energy with original starting sequence" << endl;
                cout << "runmode = 12 | Optimize binding energy with original starting sequence" << endl;
                cout << "runmode = 13 | Optimize both system and binding energies with original starting
sequence" << endl;
                exit(1);
        }


        int runmode = atoi(argv[1]);
        string infile = argv[2];
        string localout = argv[3];
        int debug = 0;
        int startrunmode = runmode;


        if (runmode < 10) {
                randStart = 1;
        }


        if (runmode > 9) {
                runmode = startrunmode - 10;
                randStart = 0;
        }


        if (argc > 4) {
                string debugOption = argv[4];
                if (debugOption == "--debug") {
                        debug = 1;
                }
        }


        if (debug == 1) {
                cout << "Debugging enabled" << endl;
        }


        enum aminoAcid
{A,R,N,D,Dh,C,Cx,Cf,Q,E,Eh,Hd,He,Hn,Hp,I,L,K,M,F,P,O,S,T,W,Y,V,G,dA,dR,dN,dD,dDh,dC,dCx,dQ,dE,dEh,dHd,dHe,dHn,dH
p,dI,dL,dK,dM,dF,dP,dO,dS,dT,dAT,dW,dY,dV,Hce,Pch,Csf};
```

```
	string aminoAcidString[] =
{"A","R","N","D","Dh","C","Cx","Cf","Q","E","Eh","Hd","He","Hn","Hp","I","L","K","M","F","P","O","S","T","W","Y"
,"V","G","dA","dR","dN","dD","dDh","dC","dCx","dQ","dE","dEh","dHd","dHe","dHn","dHp","dI","dL","dK","dM","dF","
dP","dO","dS","dT","dAT","dW","dY","dV","Hce","Pch","Csf"};

	PDBInterface* thePDB = new PDBInterface(infile);

	ensemble* theEnsemble = thePDB->getEnsemblePointer();

	molecule* pMol = theEnsemble->getMoleculePointer(0);

	protein* bundle = static_cast<protein*>(pMol);

	bundle->silenceMessages();

	residue::setCutoffDistance(9.0);

	rotamer::setScaleFactor(0.0);

	amberVDW::setScaleFactor(1.0);

	amberVDW::setRadiusScaleFactor(1.0);

	amberVDW::setLinearRepulsionDampeningOff();

	amberElec::setScaleFactor(1.0);

	srand (getpid());


	//--inputs for mutation

	UInt activeChains[] = {1};

	UInt allowedLResidues[] = {A,R,N,D,Q,E,I,L,K,M,F,S,T,W,Y,V};

	UInt activeResidues[] = {1,2,3,4,5,6,7,8,9,10};

	UInt randomResidues[] = {1,2,3,4,5,6,7,8,9,10};

	UInt allowedDResidues[] = {G};


	double phi, initialSystemEnergy, initialBindingEnergy, currentSystemEnergy, currentBindingEnergy,
bestSystemEnergy, bestBindingEnergy;

	UInt nobetter = 0, activeChainsSize = sizeof(activeChains)/sizeof(activeChains[0]), randomResiduesSize
= sizeof(randomResidues)/sizeof(randomResidues[0]), activeResiduesSize =
sizeof(activeResidues)/sizeof(activeResidues[0]);

	UInt lResidues = sizeof(allowedLResidues)/sizeof(allowedLResidues[0]), dResidues =
sizeof(allowedDResidues)/sizeof(allowedDResidues[0]);

	UInt name, timeid, sec, mutant = 0, numResidues, plateau = (0.5 * activeResiduesSize);

	stringstream convert;

	string startstr, outFile;

	name = rand() % 100000000;

	convert << name, startstr = convert.str();

	string tempModel = startstr + "_temp.pdb";

	delete thePDB;


	if (debug == 1) {

		cout << "Run Mode: " << runmode << " | Initial Mutation Mode: " << startrunmode << endl;

	}
```

```
	//--Run multiple independent evolution cycles-------------------------------------------------------
	for (UInt a = 1; a < 500; a++) {
		sec = time(NULL);
		timeid = sec;
		UInt familystr = timeid;

		if (debug == 1) {
			cout << "Evolution Cycle: " << a << " | Family ID: " << familystr << endl;
		}

		PDBInterface* thePDB = new PDBInterface(infile);
		ensemble* theEnsemble = thePDB->getEnsemblePointer();
		molecule* pMol = theEnsemble->getMoleculePointer(0);
		protein* bundle = static_cast<protein*>(pMol);
		vector < UInt > mutantPosition, chainSequence, midChainSequence, sequencePosition,
randomPosition;
		vector < vector < UInt > > sequencePool, proteinSequence, midSequence;
		sequencePool = buildSequencePool();

		vector <double> initialEnergy = bundle->chainBindingEnergy();
		initialSystemEnergy = initialEnergy[0];
		double iSE = initialSystemEnergy;
		initialBindingEnergy = initialEnergy[1];
		double iBE = initialBindingEnergy;

		if (initialSystemEnergy > 0) {
			initialSystemEnergy = (0.5 * iSE); //initialSystemEnergy becomes Target
		}

		if (initialBindingEnergy > 0) {
			initialBindingEnergy = (0.5 * iBE); //initialBindingEnergy becomes Target
		}

		//--load in initial pdb and mutate in random starting sequence on active chains and random
residues
		nobetter = 0;

				if (debug == 1) {
				    cout << "Initial random mutation starting" << endl;
				    tInitialMut = getCurrentTime();
				}
```

```
                for (UInt i = 0; i < activeChainsSize; i++) {


                        for (UInt j = 0; j < randomResiduesSize; j++) {
                              bundle->setMoved(activeChains[i],randomResidues[j],1);
                                 bundle->activateForRepacking(activeChains[i], randomResidues[j]);
                                 randomPosition.push_back(activeChains[i]);
                                 randomPosition.push_back(randomResidues[j]);
                                 phi = bundle->getPhi(activeChains[i], randomResidues[j]);


                                 if (randStart == 1) {


                                          if (phi > 0 && phi < 180) {
                                                   mutant = getProbabilisticMutation(sequencePool,
randomPosition, allowedDResidues, dResidues);
                                                   bundle->mutateWBC(activeChains[i], randomResidues[j],
mutant);
                                          }


                                          if (phi < 0 && phi > -180) {
                                                   mutant = getProbabilisticMutation(sequencePool,
randomPosition, allowedLResidues, lResidues);
                                                   bundle->mutateWBC(activeChains[i], randomResidues[j],
mutant);
                                          }


                                 }


                                 randomPosition.clear();
                        }
                        chainSequence = getChainSequence(bundle, activeChains[i]);
                        proteinSequence.push_back(chainSequence);
                }
                if (debug == 1) {
                         tPostInitialMut = getCurrentTime() - tInitialMut;
                        cout << "Initial random mutation created in " << tPostInitialMut << " seconds,
optimizing system" << endl;
                tInitialOpt = getCurrentTime();
                }


                bundle->protOpt(false);
```

```
                    if (debug == 1) {

                              tPostInitialOpt = getCurrentTime() - tInitialOpt;

                              cout << "Initial mutant optimized in " << tPostInitialOpt << " seconds" << endl;

                    }


                    cout << "Starting Energies | System Energy: " << iSE << " | Initial System Energy: " <<
initialSystemEnergy << " | Binding Energy: " << iBE << " | Initial Binding Energy: " << initialBindingEnergy <<
endl;



                    //--Determine next mutation position

                    mutantPosition.clear();

                    mutantPosition = getMutationPosition(bundle, activeChains, activeChainsSize, activeResidues,
activeResiduesSize);

                    pdbWriter(bundle, tempModel);


                    //--set Energy startpoint

                    vector <double> totalEnergy = bundle->chainBindingEnergy();

                    currentSystemEnergy = totalEnergy[0];

                    currentBindingEnergy = totalEnergy[1];


                    //        if ((runmode < 2) && (currentSystemEnergy > 0))

                    //        {currentSystemEnergy = 0;

                    //        }


                    //        if ((runmode > 1) && (currentBindingEnergy > 0))

                    //        {currentBindingEnergy = 0;

                    //        }


                    bestSystemEnergy = currentSystemEnergy;


                    bestBindingEnergy = currentBindingEnergy;


                    if (debug == 1) {

                    cout << "Starting Energies | Target System Energy: " << initialSystemEnergy << " | Target
Binding Energy: " << initialBindingEnergy << " | Current System Energy: " << currentSystemEnergy << " | Current
Binding Energy: " << currentBindingEnergy << endl;

                    }


                    delete thePDB;


                    UInt finalPlateau = (plateau + runmode);
```

```
		//--Run through a single evolutionary path (ancestral line) till hitting plateau
		do {

			if (debug == 1) {
				cout << "Do loop started" << endl;
			}

			PDBInterface* thePDB = new PDBInterface(infile);
			ensemble* theEnsemble = thePDB->getEnsemblePointer();
			molecule* pMol = theEnsemble->getMoleculePointer(0);
			protein* bundle = static_cast<protein*>(pMol);
			//--Mutate current sequence, new mutant and optimize system
			nobetter++;

			if (debug == 1) {
				cout << "Mutation starting" << endl;
				tMutStart = getCurrentTime();
			}

			for (UInt i = 0; i < activeChainsSize; i++) {
				numResidues = bundle->getNumResidues(activeChains[i]);

				for (UInt j = 0; j < numResidues; j++) {
					bundle->setMoved(activeChains[i], j, 1);
					bundle->activateForRepacking(activeChains[i],j);

					if (activeChains[i] == mutantPosition[0] && j ==
mutantPosition[1]) {
						//--new mutant
						sequencePosition.push_back(i);
						sequencePosition.push_back(j);
						phi = bundle->getPhi(mutantPosition[0],
mutantPosition[1]);

						if (phi > 0 && phi < 180) {
							mutant =
getProbabilisticMutation(sequencePool, mutantPosition, allowedDResidues, dResidues);
							bundle-
>mutateWBC(mutantPosition[0],mutantPosition[1], mutant);
						}
```

```
                                                  if (phi < 0 && phi > -180) {
                                                            mutant =
getProbabilisticMutation(sequencePool, mutantPosition, allowedLResidues, lResidues);
                                                            bundle-
>mutateWBC(mutantPosition[0],mutantPosition[1], mutant);
                                                  }

                                        }
                                        else{
                                                  bundle->mutateWBC(activeChains[i],j,
proteinSequence[i][j]);
                                        }
                              }
                              //randomizeSideChains(bundle, activeChains[i]);
                    }

                    if (debug == 1) {
                              tMutFinished = getCurrentTime() - tMutStart;
                              cout << "Mutation created in " << tMutFinished << " seconds, optimizing
system" << endl;
                              tMutOptStart = getCurrentTime();
                    }




                    bundle->protOpt(false);



                    if (debug == 1) {
                              tMutOptFinished = getCurrentTime() - tMutOptStart;
                              cout << "Mutant optimized in " << tMutOptFinished << " seconds" << endl;
                    }

                    protein* tempBundle = new protein(*bundle);

                    if (debug == 1) {
                              cout << "New tempBundle generated" << endl;
                    }

                    //--Determine next mutation position
                    mutantPosition.clear();
```

```
                        mutantPosition = getMutationPosition(bundle, activeChains, activeChainsSize,
activeResidues, activeResiduesSize);


                        if (debug == 1) {
                                cout << "New mutation position determined" << endl;
                        }


                        //--Energy test
                        vector <double> totalEnergy = bundle->chainBindingEnergy();
                        currentSystemEnergy = totalEnergy[0];
                        currentBindingEnergy = totalEnergy[1];


                        double targetSystemEnergy = initialSystemEnergy;
                        double targetBindingEnergy = initialBindingEnergy;


                        if (debug == 1) {
                                cout << "Energies updated" << endl;
                        }



                        // if System Energy is positive, target for at least 20% energy removal per loop
                        if (currentSystemEnergy > 0) {
                                targetSystemEnergy = bestSystemEnergy * 0.8;
                        }


                        // if Binding Energy is positive, target for at least 20% energy removal per loop
                        if (currentBindingEnergy > 0) {
                                targetBindingEnergy = bestBindingEnergy * 0.8;
                        }


                        // if System Energy is negative, target for at least 1% energy removal per loop
                        if ((currentSystemEnergy < 0) || (currentSystemEnergy == 0)) {
                                targetSystemEnergy = bestSystemEnergy * 1.01;
                        }


                        // if Binding Energy is positive, target for at least 1% energy removal per loop
                        if ((currentBindingEnergy < 0) || (currentBindingEnergy == 0)) {
                                targetBindingEnergy = bestBindingEnergy * 1.01;
                        }
```

```
                        if (debug == 1) {

                                cout << "Starting Energies | Target System Energy: " << targetSystemEnergy
<< " | Target Binding Energy: " << targetBindingEnergy << " | Current System Energy: " << currentSystemEnergy <<
" | Current Binding Energy: " << currentBindingEnergy << " | Plateau: " << finalPlateau << " | Buffer: " <<
nobetter << endl;

                        }


                        if ((runmode < 2) && (currentSystemEnergy < targetSystemEnergy)) {

                        //cout << Energy << " " << nobetter << endl;




                                if ((currentSystemEnergy < bestSystemEnergy) && (currentSystemEnergy <
targetSystemEnergy)) {

                                        bestSystemEnergy = currentSystemEnergy;

                                        pdbWriter(tempBundle, tempModel);


                                        proteinSequence[sequencePosition[0]][sequencePosition[1]] =
mutant;


                                        if (((runmode < 2) && (currentSystemEnergy <
initialSystemEnergy)) || ((runmode == 2) && (currentBindingEnergy < initialBindingEnergy) &&
(currentSystemEnergy < 0) && (currentBindingEnergy < 0))         || ((runmode == 3) && (currentSystemEnergy <
initialSystemEnergy) && (currentBindingEnergy < initialBindingEnergy) && (currentBindingEnergy < 0))) {


                                                if (debug == 1) {

                                                        cout << "Found a suitable sequence, printing
PDB" << endl;

                                                }


                                                PDBInterface* theModelPDB = new PDBInterface(tempModel);

                                                ensemble* theModelEnsemble = theModelPDB-
>getEnsemblePointer();

                                                molecule* modelMol = theModelEnsemble-
>getMoleculePointer(0);

                                                protein* model = static_cast<protein*>(modelMol);


                                                name = rand() % 100;

                                                sec = time(NULL);

                                                timeid = name + sec;

                                                stringstream convert;

                                                string countstr;

                                                convert << timeid, countstr = convert.str();

                                                outFile = countstr + ".evo.pdb";

                                                pdbWriter(model, outFile);
```

```
activeChains[i]);

| fstream::app);

<< currentBindingEnergy << " ";

fstream::app);

" << currentBindingEnergy << " ";

fstream::out | fstream::app);

proteinSequence[i].size(); j++) {

aminoAcidString[proteinSequence[i][j]] << " ";

aminoAcidString[proteinSequence[i][j]] << " ";
```

```
midSequence.clear(), midChainSequence.clear();

for (UInt i = 0; i < activeChainsSize; i++) {
          midChainSequence = getChainSequence(model,
          midSequence.push_back(chainSequence);
}

fstream finalline;
finalline.open ("final.out", fstream::in | fstream::out
finalline << timeid << " " << currentSystemEnergy << " "

fstream finallocal;
finallocal.open (localout, fstream::in | fstream::out |
finallocal << timeid << " " << currentSystemEnergy << "

fstream fs;
fs.open ("finalsequences.out", fstream::in |

for (UInt i = 0; i < proteinSequence.size(); i++) {

          for (UInt j = 0; j <
                    finalline <<
                    finallocal <<
                    fs << proteinSequence[i][j] << ",";
          }

}

fs << endl;
finalline << " " << familystr << " " << endl;
finalline.close();
finallocal << " " << familystr << " " << endl;
finallocal.close();
fs.close();
delete theModelPDB;
```

```
                                        }

                                        if (debug == 1) {
                                                cout << "Temp updated, System Energy Better" << endl;
                                        }
                                        nobetter = 0;
                                }

                                if (nobetter > 0) {
                                        nobetter--;
                                }

                                else{
                                        nobetter = 0;
                                }
                        }

                        if (runmode == 2 && (currentBindingEnergy < targetBindingEnergy)) {
                                //cout << Energy << " " << nobetter << endl;
                                if ((currentBindingEnergy < bestBindingEnergy) && (currentBindingEnergy <
targetBindingEnergy)) {
                                        bestBindingEnergy = currentBindingEnergy;
                                        pdbWriter(tempBundle, tempModel);
                                        proteinSequence[sequencePosition[0]][sequencePosition[1]] =
mutant;

                                        if (debug == 1) {
                                                cout << "Temp updated, Binding Energy Better" << endl;
                                        }

                                        nobetter = 0;
                                }

                                if ((runmode == 2) && (currentBindingEnergy < ((initialBindingEnergy +
targetBindingEnergy) * 0.5)) && (currentSystemEnergy < 0) && (currentBindingEnergy < 0)) {

                                        if (debug == 1) {
                                                cout << "Found a suitable sequence, printing PDB" <<
endl;
                                        }

                                        PDBInterface* theModelPDB = new PDBInterface(tempModel);
```

```
ensemble* theModelEnsemble = theModelPDB->getEnsemblePointer();
molecule* modelMol = theModelEnsemble->getMoleculePointer(0);
protein* model = static_cast<protein*>(modelMol);

name = rand() % 100;
sec = time(NULL);
timeid = name + sec;
stringstream convert;
string countstr;
convert << timeid, countstr = convert.str();
outFile = countstr + ".evo.pdb";
pdbWriter(model, outFile);
midSequence.clear(), midChainSequence.clear();

for (UInt i = 0; i < activeChainsSize; i++) {
          midChainSequence = getChainSequence(model,
activeChains[i]);
          midSequence.push_back(chainSequence);
}

fstream finalline;
finalline.open ("final.out", fstream::in | fstream::out |
fstream::app);
finalline << timeid << " " << currentSystemEnergy << " " <<
currentBindingEnergy << " ";

fstream finallocal;
finallocal.open (localout, fstream::in | fstream::out |
fstream::app);
finallocal << timeid << " " << currentSystemEnergy << " " <<
currentBindingEnergy << " ";

fstream fs;
fs.open ("finalsequences.out", fstream::in | fstream::out |
fstream::app);

for (UInt i = 0; i < proteinSequence.size(); i++) {

          for (UInt j = 0; j < proteinSequence[i].size(); j++) {
                    finalline <<
aminoAcidString[proteinSequence[i][j]] << " ";
                    finallocal <<
aminoAcidString[proteinSequence[i][j]] << " ";
                    fs << proteinSequence[i][j] << ",";
```

```
                                        }


                                }
                                fs << endl;
                                finalline << " " << familystr << " " << endl;
                                finalline.close();
                                finallocal << " " << familystr << " " << endl;
                                finallocal.close();
                                fs.close();
                                delete theModelPDB;


                        }
                        if (nobetter > 0) {
                                nobetter--;
                        }
                        else{
                                nobetter = 0;
                        }
                }


                        if ((runmode == 3)  && (((currentSystemEnergy < bestSystemEnergy) &&
(currentBindingEnergy < initialBindingEnergy)) || ((currentSystemEnergy < initialSystemEnergy) &&
(currentBindingEnergy < bestBindingEnergy)))) {



                        //cout << Energy << " " << nobetter << endl;
                                if (((currentSystemEnergy < bestSystemEnergy) && (currentBindingEnergy <
targetBindingEnergy)) || ((currentSystemEnergy < targetSystemEnergy) && (currentBindingEnergy <
bestBindingEnergy))) {



                                        if (currentSystemEnergy < bestSystemEnergy) {
                                                bestSystemEnergy = currentSystemEnergy;
                                        }


                                        if (currentBindingEnergy < bestBindingEnergy) {
                                                bestBindingEnergy = currentBindingEnergy;
                                        }


                                        pdbWriter(tempBundle, tempModel);


                                        proteinSequence[sequencePosition[0]][sequencePosition[1]] =
mutant;
```

```
                                        if (((runmode < 2) && (currentSystemEnergy <
initialSystemEnergy)) || ((runmode == 2) && (currentBindingEnergy < initialBindingEnergy) &&
(currentSystemEnergy < 0) && (currentBindingEnergy < 0)) || ((runmode == 3) && (currentSystemEnergy <
initialSystemEnergy) && (currentBindingEnergy < initialBindingEnergy) && (currentBindingEnergy < 0))) {


                                                if (debug == 1) {
                                                        cout << "Found a suitable sequence, printing
PDB" << endl;
                                                }


                                                PDBInterface* theModelPDB = new PDBInterface(tempModel);
                                                ensemble* theModelEnsemble = theModelPDB-
>getEnsemblePointer();
                                                molecule* modelMol = theModelEnsemble-
>getMoleculePointer(0);
                                                protein* model = static_cast<protein*>(modelMol);


                                                name = rand() % 100;
                                                sec = time(NULL);
                                                timeid = name + sec;
                                                stringstream convert;
                                                string countstr;
                                                convert << timeid, countstr = convert.str();
                                                outFile = countstr + ".evo.pdb";
                                                pdbWriter(model, outFile);
                                                midSequence.clear(), midChainSequence.clear();


                                                for (UInt i = 0; i < activeChainsSize; i++) {
                                                        midChainSequence = getChainSequence(model,
activeChains[i]);
                                                        midSequence.push_back(chainSequence);
                                                }


                                                fstream finalline;
                                                finalline.open ("final.out", fstream::in | fstream::out
| fstream::app);
                                                finalline << timeid << " " << currentSystemEnergy << " "
<< currentBindingEnergy << " ";


                                                fstream finallocal;
                                                finallocal.open (localout, fstream::in | fstream::out |
fstream::app);
                                                finallocal << timeid << " " << currentSystemEnergy << "
" << currentBindingEnergy << " ";
```

```
                                                fstream fs;
                                                fs.open ("finalsequences.out", fstream::in |
fstream::out | fstream::app);

                                                for (UInt i = 0; i < proteinSequence.size(); i++) {

                                                        for (UInt j = 0; j <
proteinSequence[i].size(); j++) {
                                                                finalline <<
aminoAcidString[proteinSequence[i][j]] << " ";
                                                                finallocal <<
aminoAcidString[proteinSequence[i][j]] << " ";
                                                                fs << proteinSequence[i][j] << ",";
                                                        }
                                                }
                                                fs << endl;
                                                finalline << " " << familystr << " " << endl;
                                                finalline.close();
                                                finallocal << " " << familystr << " " << endl;
                                                finallocal.close();
                                                fs.close();

                                                delete theModelPDB;

                                        }

                                        if (debug == 1) {
                                                cout << "Temp updated, System Energy Better" << endl;
                                        }

                                nobetter = 0;
                                }

                                if (nobetter > 0) {
                                        nobetter--;
                                }
                                else{
                                        nobetter = 0;
                                }
                        }

                        sequencePosition.clear();
```

```
                        delete thePDB;

                        delete tempBundle;

                } while (nobetter < finalPlateau);


        //--Print final energy and write a pdb file-----------------------------------------------------


        sequencePool.clear(),proteinSequence.clear(), chainSequence.clear(), mutantPosition.clear(),
chainSequence.clear(), sequencePosition.clear(), randomPosition.clear();

        }

        return 0;

}




////////////////////////////////////////////////////////////////////////////////////////////////////

//////// functions ////////////////////////////////////////////////////////////////////////////////

////////////////////////////////////////////////////////////////////////////////////////////////////


void randomizeSideChains(protein* _prot, UInt _chainIndex) {

    UInt allowedRotsSize, randrot, restype, numResidues;

    UIntVec allowedRots;

    numResidues = _prot->getNumResidues(_chainIndex);


        for (UInt j = 0; j < numResidues; j++) {

                restype = _prot->getTypeFromResNum(_chainIndex, j);

                allowedRots = _prot->getAllowedRotamers(_chainIndex, j, restype, 0);

                allowedRotsSize = allowedRots.size();


                if (allowedRotsSize > 2) {

                        randrot = rand() % allowedRotsSize;

                        _prot->setRotamerWBC(_chainIndex, j, 0, allowedRots[randrot]);

                }


        }


    return;

}


vector < UInt > getChainSequence(protein* _prot, UInt _chainIndex) {

        UInt restype, numResidues;

        vector < UInt > sequence;
```

```
	numResidues = _prot->getNumResidues(_chainIndex);


	for (UInt j = 0; j < numResidues; j++) {
		restype = _prot->getTypeFromResNum(_chainIndex, j);
		sequence.push_back(restype);
	}


	return sequence;
}


vector <UInt> getMutationPosition(protein* _prot, UInt *_activeChains, UInt _activeChainsSize, UInt
*_activeResidues, UInt _activeResiduesSize) {


	UIntVec activeChains, activeResidues;
	    for (UInt i = 0; i < _activeChainsSize; i++)
	    {
		activeChains.push_back(_activeChains[i]);
	    }
	    for (UInt i = 0; i < _activeResiduesSize; i++)
	    {
		activeResidues.push_back(_activeResidues[i]);
	    }


	//--get median residue energy
	UInt randres, randchain;
	double posE, medE;
	vector <UInt> _mutantPosition;
	medE = _prot->getMedianResEnergy(activeChains, activeResidues);


	//--find random position with worse than median energy
	do {
		randchain = _activeChains[rand() % _activeChainsSize];
		randres = _activeResidues[rand() % _activeResiduesSize];
		posE = _prot->resEnergy(randchain,randres);
	} while (posE < medE);
		_mutantPosition.push_back(randchain);
		_mutantPosition.push_back(randres);
		return _mutantPosition;
}
```

```
UInt getProbabilisticMutation(vector < vector < UInt > > &_sequencePool, vector <UInt> _mutantPosition, UInt
*_aminoacids, UInt aaSize) {

	double mutant, chance, entropy;

	double acceptance;

	vector <UInt> resFreqs(58,1);

	double count = _sequencePool.size();



	//--get sequence evolution results for position

	for (UInt i = 0; i < _sequencePool.size(); i++) {

		UInt restype = _sequencePool[i][_mutantPosition[1]];

		resFreqs[restype] = resFreqs[restype] + 1;

	}



	//--determine population based chance of mutation acceptance or a random mutation, via linear
regression of sequence entropy

	do {

		chance = rand() % 100;

		entropy = rand() % 100; //sequence entropy determined by pooling linear decline to resolve
minima after suitable diversity

		mutant = _aminoacids[rand() % aaSize];

		double pooling = -0.09 * count + 140; //500 sequences equals 5% chance of pooling sequences,
95% chance of it being random, 1000=50% ...



		if (entropy > pooling) {

			acceptance = (resFreqs[mutant]/count)*100; //chance of accepting given amino acid at
position is proportional to population

		}



		else{

			acceptance = 100;  //random mutation

		}



	} while (chance > acceptance);

		return mutant;

}



vector < vector < UInt > > buildSequencePool() {

	ifstream file("finalsequences.out");

	string item, line;

	vector < UInt > sequence;

	vector < vector < UInt > > sequencePool;
```

```
        while(getline(file,line)) {
                stringstream stream(line);

                while(getline(stream,item,',')) {
                        stringstream aaString(item);
                        int aaIndex;
                        aaString >> aaIndex;
                        sequence.push_back(aaIndex);
                }

        sequencePool.push_back(sequence);
        sequence.clear();
        }

        file.close();
        return sequencePool;
}

int getCurrentTime() {
    time_t now = time(0);
    return now;
}
```

BIBLIOGRAPHY

1. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302,** 1364–8 (2003).

2. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. in *Methods in Enzymology* **383,** 66–93 (2004).

3. MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. *J. Phys. Chem. B* **102,** 3586–3616 (1998).

4. Unger, R. & Moult, J. Finding the lowest free energy conformation of a protein is an NP-hard problem: Proof and implications. *Bull. Math. Biol.* **55,** 1183–1198 (1993).

5. Desmet, J., Maeyer, M. De, Hazes, B. & Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **356,** 539–542 (1992).

6. Chen, M. & Huang, W. Q. A branch and bound algorithm for the protein folding problem in the HP lattice model. *Genomics. Proteomics Bioinformatics* **3,** 225–30 (2005).

7. Pokala, N. & Handel, T. M. Review: Protein Design—Where We Were, Where We Are, Where We're Going. *J. Struct. Biol.* **134,** 269–281 (2001).

8. Moult, J., Pedersen, J. T., Judson, R. & Fidelis, K. A large-scale experiment to assess protein structure prediction methods. *Proteins Struct. Funct. Genet.* **23,** ii–iv (1995).

9. Ovchinnikov, S. *et al.* Improved de novo structure prediction in CASP11 by incorporating Co-evolution information into rosetta. *Proteins Struct. Funct. Bioinforma.* n/a–n/a (2015). doi:10.1002/prot.24974

10. Lam, H. *et al.* D-Amino Acids Govern Stationary Phase Cell Wall Remodeling in Bacteria. *Science (80-.).* **325,** 1552–1555 (2009).

11. Chervyakov, A. V., Gulyaeva, N. V. & Zakharova, M. N. D-amino acids in normal ageing and pathogenesis of neurodegenerative diseases. *Neurochem. J.* **5,** 100–114 (2011).

12. Fujii, N. D-Amino Acid in Elderly Tissues. *Biol. Pharm. Bull.* **28,** 1585–1589 (2005).

13. Haack, T., González, M. J., Sánchez, Y. & Giralt, E. D-Amino acids in protein de novo design. II. Protein-diastereomerism versus protein-enantiomerism. *Lett. Pept. Sci.* **4,** 377–386 (1997).

14. Jeffrey, G. A. *An Introduction to Hydrogen Bonding*. (Oxford University Press, 1997).

15. Aurora, R. & Rosee, G. D. Helix capping. *Protein Sci.* **7,** 21–38 (1998).

16. Sagermann, M., Mårtensson, L.-G., Baase, W. A. & Matthews, B. W. A test of proposed rules for helix capping: Implications for protein design. *Protein Sci.* **11,** 516–521 (2009).

17. Kapp, G. T., Richardson, J. S. & Oas, T. G. Kinetic Role of Helix Caps in Protein Folding Is Context-Dependent †. *Biochemistry* **43,** 3814–3823 (2004).

18. Anil, B., Song, B., Tang, Y. & Raleigh, D. P. Exploiting the Right Side of the Ramachandran Plot: Substitution of Glycines by <scp>d</scp> -Alanine Can Significantly Increase Protein Stability. *J. Am. Chem. Soc.* **126,** 13194–13195 (2004).

19. Bang, D. *et al.* Dissecting the energetics of protein α-helix C-cap termination through chemical protein synthesis. *Nat. Chem. Biol.* **2,** 139–143 (2006).

20. Rodriguez-Granillo, A., Annavarapu, S., Zhang, L., Koder, R. L. & Nanda, V. Computational Design of Thermostabilizing D-Amino Acid Substitutions. *J. Am. Chem. Soc.* **133,** 18750–18759 (2011).

21. Annavarapu, S. & Nanda, V. Mirrors in the PDB: left-handed α-turns guide design with D-amino acids. *BMC Struct. Biol.* **9,** 61 (2009).

22. Haile, J. M. *Molecular Dynamics Simulation: Elementary Methods*. (Wiley Interscience, 1997).

23. Xia, B., Tsui, V., Case, D. A., Dyson, H. J. & Wright, P. E. Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J. Biomol. NMR* **22,** 317–331 (2002).

24. Roux, B. & Simonson, T. Implicit solvent models. *Biophys. Chem.* **78,** 1–20 (1999).

25. Feig, M. & Brooks, C. L. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **14,**

217–224 (2004).

26. Alder, B. J. & Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **31,** 459 (1959).

27. Qiu, L., Pabit, S. A., Roitberg, A. E. & Hagen, S. J. Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 μs. *J. Am. Chem. Soc.* **124,** 12952–12953 (2002).

28. Götz, A. W. *et al.* Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **8,** 1542–1555 (2012).

29. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10,** 449–461 (2015).

30. Miller, B. R. *et al.* MMPBSA.py : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **8,** 3314–3321 (2012).

31. Stein, E. G., Rice, L. M. & Brünger, A. T. Torsion-Angle Molecular Dynamics as a New Efficient Tool for NMR Structure Calculation. *J. Magn. Reson.* **124,** 154–164 (1997).

32. Mason, J. M. Design and development of peptides and peptide mimetics as antagonists for therapeutic intervention. *Future Med. Chem.* **2,** 1813–1822 (2010).

33. Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* **450,** 1001–1009 (2007).

34. Lindgren, M., Hällbrink, M., Prochiantz, A. & Langel, Ü. Cell-penetrating

peptides. *Trends Pharmacol. Sci.* **21,** 99–103 (2000).

35. Baker, M., Reynolds, H. M., Lumicisi, B. & Bryson, C. J. Immunogenicity of protein therapeutics: The key causes, consequences and challenges. *Self. Nonself.* **1,** 314–322 (2010).

36. Fosgerau, K. & Hoffmann, T. Peptide therapeutics: current status and future directions. *Drug Discov. Today* **20,** 122–128 (2015).

37. Pernot, M., Vanderesse, R., Frochot, C., Guillemin, F. & Barberi-Heyob, M. Stability of peptides and therapeutic success in cancer. *Expert Opin. Drug Metab. Toxicol.* **7,** 793–802 (2011).

38. Nauck, M. A., Holst, J. J., Willms, B. & Schmiegel, W. Glucagon-like peptide 1 (GLP-1) as a new therapeutic approach for Type 2-diabetes. *Exp. Clin. Endocrinol. & Diabetes* **105,** 187–195 (1997).

39. Barnhart, K. F. *et al.* A Peptidomimetic Targeting White Fat Causes Weight Loss and Improved Insulin Resistance in Obese Monkeys. *Sci. Transl. Med.* **3,** 108ra112–108ra112 (2011).

40. Welch, B. D., VanDemark, A. P., Heroux, A., Hill, C. P. & Kay, M. S. Potent D-peptide inhibitors of HIV-1 entry. *Proc. Natl. Acad. Sci.* **104,** 16828–16833 (2007).

41. Walensky, L. D. & Bird, G. H. Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress. *J. Med. Chem.* **57,** 6275–6288 (2014).

42. Stewart, M. L., Fire, E., Keating, A. E. & Walensky, L. D. The MCL-1 BH3 helix

is an exclusive MCL-1 inhibitor and apoptosis sensitizer. *Nat. Chem. Biol.* **6,** 595–601 (2010).

43. Dushoff, J. Mortality due to Influenza in the United States--An Annualized Regression Approach Using Multiple-Cause Mortality Data. *Am. J. Epidemiol.* **163,** 181–187 (2005).

44. Taubenberger, J. K. & Morens, D. M. 1918 Influenza: the Mother of All Pandemics. *Emerg. Infect. Dis.* **12,** 15–22 (2006).

45. Osterholm, M. T., Kelley, N. S., Sommer, A. & Belongia, E. A. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. *Lancet Infect. Dis.* **12,** 36–44 (2012).

46. Hayden, F. G. & de Jong, M. D. Emerging Influenza Antiviral Resistance Threats. *J. Infect. Dis.* **203,** 6–10 (2011).

47. Poland, G. A., Jacobson, R. M. & Ovsyannikova, I. G. Influenza Virus Resistance to Antiviral Agents: A Plea for Rational Use. *Clin. Infect. Dis.* **48,** 1254–1256 (2009).

48. Khan, F. H. *The Elements of Immunology*. (Prentice Hall, 2009).

49. Rust, M. J., Lakadamyali, M., Zhang, F. & Zhuang, X. Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nat. Struct. Mol. Biol.* **11,** 567–573 (2004).

50. Samji, T. Influenza A: understanding the viral life cycle. *Yale J. Biol. Med.* **82,** 153–9 (2009).

51. Matsuoka, Y. *et al.* A comprehensive map of the influenza A virus replication cycle. *BMC Syst. Biol.* **7,** 97 (2013).

52. Bouloy, M., Morgan, M., Shatkin, A. & Krug, R. Cap and internal nucleotides of reovirus mRNA primers are incorporated into influenza viral complementary RNA during transcription in vitro. *J. Virol.* **32,** 895–904 (1979).

53. Porter, A. G., Smith, J. C. & Emtage, J. S. Nucleotide sequence of influenza virus RNA segment 8 indicates that coding regions for NS1 and NS2 proteins overlap. *Proc. Natl. Acad. Sci.* **77,** 5074–8 (1980).

54. Wolstenholme, A. J., Barrett, T., Nichol, S. T. & Mahy, B. W. Influenza virus-specific RNA and protein syntheses in cells infected with temperature-sensitive mutants defective in the genome segment encoding nonstructural proteins. *J. Virol.* **35,** 1–7 (1980).

55. Hale, B. G., Randall, R. E., Ortin, J. & Jackson, D. The multifunctional NS1 protein of influenza A viruses. *J. Gen. Virol.* **89,** 2359–2376 (2008).

56. Yin, C. *et al.* Conserved Surface Features Form the Double-stranded RNA Binding Site of Non-structural Protein 1 (NS1) from Influenza A and B Viruses. *J. Biol. Chem.* **282,** 20584–20592 (2007).

57. Chien, C. *et al.* A novel RNA-binding motif in influenza A virus non-structural protein 1. *Nat. Struct. Biol.* **4,** 891–895 (1997).

58. Bergmann, M. *et al.* Influenza Virus NS1 Protein Counteracts PKR-Mediated Inhibition of Replication. *J. Virol.* **74,** 6203–6206 (2000).

59. Aramini, J. M. *et al.* 19F NMR Reveals Multiple Conformations at the Dimer Interface of the Nonstructural Protein 1 Effector Domain from Influenza A Virus. *Structure* **22,** 515–525 (2014).

60. Das, K. *et al.* Structural basis for suppression of a host antiviral response by influenza A virus. *Proc. Natl. Acad. Sci.* **105,** 13093–13098 (2008).

61. Nemeroff, M. E., Barabino, S. M. L., Li, Y., Keller, W. & Krug, R. M. Influenza Virus NS1 Protein Interacts with the Cellular 30 kDa Subunit of CPSF and Inhibits 3′ End Formation of Cellular Pre-mRNAs. *Mol. Cell* **1,** 991–1000 (1998).

62. Wang, X. *et al.* The non-structural (NS1) protein of influenza A virus associates with p53 and inhibits p53-mediated transcriptional activity and apoptosis. *Biochem. Biophys. Res. Commun.* **395,** 141–145 (2010).

63. Murayama, R. *et al.* Influenza A virus non-structural protein 1 (NS1) interacts with cellular multifunctional protein nucleolin during infection. *Biochem. Biophys. Res. Commun.* **362,** 880–885 (2007).

64. Ngamurulert, S., Limjindaporn, T. & Auewaraku, P. Identification of cellular partners of Influenza A virus (H5N1) non-structural protein NS1 by yeast two-hybrid system. *Acta Virol.* **53,** 153–159 (2009).

65. Aramini, J. M. *et al.* Dimer Interface of the Effector Domain of Non-structural Protein 1 from Influenza A Virus: AN INTERFACE WITH MULTIPLE FUNCTIONS. *J. Biol. Chem.* **286,** 26050–26060 (2011).

66. Carrillo, B. *et al.* The Influenza A Virus Protein NS1 Displays Structural

Polymorphism. *J. Virol.* **88,** 4113–4122 (2014).

67. Montelione, G. T., Zheng, D., Huang, Y. J., Gunsalus, K. C. & Szyperski, T. Protein NMR spectroscopy in structural genomics. *Nat. Struct. Biol.* **7 Suppl,** 982–5 (2000).

68. Marsh, E. N. G. & Suzuki, Y. Using 19 F NMR to Probe Biological Interactions of Proteins and Peptides. *ACS Chem. Biol.* **9,** 1242–1250 (2014).

69. Arntson, K. E. & Pomerantz, W. C. K. Protein-Observed Fluorine NMR: A Bioorthogonal Approach for Small Molecule Discovery. *J. Med. Chem.* acs.jmedchem.5b01447 (2015). doi:10.1021/acs.jmedchem.5b01447

70. Illinois, U. of. Modeling Biological Processes at Hybrid Resolutions. (2016). Available at: http://www.ks.uiuc.edu/Research/cgfolding/.

71. Lederberg, J. H1N1-influenza as Lazarus: Genomic resurrection from the tomb of an unknown. *Proc. Natl. Acad. Sci.* **98,** 2115–2116 (2001).

72. Wernisch, L., Hery, S. & Wodak, S. J. Automatic protein design with all atom force-fields by exact and heuristic optimization. *J. Mol. Biol.* **301,** 713–736 (2000).

73. Gordon, D. B. & Mayo, S. L. Branch-and-terminate: a combinatorial optimization algorithm for protein design. *Structure* **7,** 1089–1098 (1999).

74. Kuhlman, B. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science (80-.).* **302,** 1364–1368 (2003).

75. Powell, M. F. *et al.* Peptide stability in drug development. II. Effect of single amino acid substitution and glycosylation on peptide reactivity in human serum.

Pharm. Res. **10,** 1268–73 (1993).

76. Avanti, C. *Innovative Strategies for Stabilization of Therapeutic Peptides in Aqueous Formulations*. (2012).

77. Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. The penultimate rotamer library. *Proteins Struct. Funct. Genet.* **40,** 389–408 (2000).

78. Zhou, A. Q., O'Hern, C. S. & Regan, L. The power of hard-sphere models: Explaining side-chain dihedral angle distributions of thr and val. *Biophys. J.* **102,** 2345–2352 (2012).

79. Ho, B. K. & Dill, K. a. Folding very short peptides using molecular dynamics. *PLoS Comput. Biol.* **2,** e27 (2006).

80. Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **9,** 425–430 (2002).

81. Simmerling, C., Strockbine, B. & Roitberg, A. E. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **124,** 11258–9 (2002).

82. Neidigh, J. W., Fesinmeyer, R. M., Prickett, K. S. & Andersen, N. H. Exendin-4 and Glucagon-like-peptide-1: NMR Structural Comparisons in the Solution and Micelle-Associated States †. *Biochemistry* **40,** 13188–13200 (2001).

83. Seshasayee, A. S. N. High-temperature unfolding of a trp-cage mini-protein: a molecular dynamics simulation study. *Theor. Biol. Med. Model.* **2,** 7 (2005).

84. Snow, C. D., Nguyen, H., Pande, V. S. & Gruebele, M. Absolute comparison of

simulated and experimental protein-folding dynamics. *Nature* **420,** 102–6 (2002).

85. Summa, C. protCAD. (2002).

86. Pike, D. H. & Nanda, V. Empirical estimation of local dielectric constants: Toward atomistic design of collagen mimetic peptides. *Biopolymers* **104,** 360–370 (2015).

87. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26,** 1668–88 (2005).

88. AMBER 15. (2015).

89. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).

90. Pettersen, E. F. *et al.* UCSF Chimera--A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

91. DASGUPTA, S. & BELL, J. A. Design of helix ends. *Int. J. Pept. Protein Res.* **41,** 499–511 (2009).

92. Seale, J. W., Srinivasan, R. & Rose, G. D. Sequence determinants of the capping box, a stabilizing motif at the N-termini of α-helices. *Protein Sci.* **3,** 1741–1745 (1994).

93. Harper, E. T. & Rose, G. D. Helix stop signals in proteins and peptides: The capping box. *Biochemistry* **32,** 7605–7609 (1993).

94. Spera, S. & Bax, A. Empirical correlation between protein backbone conformation and C.alpha. and C.beta. 13C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113,** 5490–5492 (1991).

95. Jia, D. *et al.* Influenza Virus Non-Structural Protein 1 (NS1) Disrupts Interferon Signaling. *PLoS One* **5,** e13927 (2010).

96. Kuo, R.-L., Zhao, C., Malur, M. & Krug, R. M. Influenza A virus strains that circulate in humans differ in the ability of their NS1 proteins to block the activation of IRF3 and interferon-β transcription. *Virology* **408,** 146–58 (2010).

97. Darapaneni, V., Prabhaker, V. K. & Kukol, A. Large-scale analysis of influenza A virus sequences reveals potential drug target sites of non-structural proteins. *J. Gen. Virol.* **90,** 2124–2133 (2009).

98. Bouvier, N. M. & Palese, P. The biology of influenza viruses. *Vaccine* **26,** D49–D53 (2008).

99. Wang, W. & Krug, R. M. The RNA-Binding and Effector Domains of the Viral NS1 Protein Are Conserved to Different Extents among Influenza A and B Viruses. *Virology* **223,** 41–50 (1996).

100. Das, K., Aramini, J. M., Ma, L.-C., Krug, R. M. & Arnold, E. Structures of influenza A proteins and insights into antiviral drug targets. *Nat. Struct. Mol. Biol.* **17,** 530–8 (2010).

101. Twu, K., Noah, D., Rao, P., Kuo, R. & Krug, R. The CPSF30 binding site on the NS1A protein of influenza A virus is a potential antiviral target. *J. Virol.* **80,** 3957–3965 (2006).

102. Aramini, J. M. *et al.* Dimer Interface of the Effector Domain of Non-structural Protein 1 from Influenza A Virus: AN INTERFACE WITH MULTIPLE

FUNCTIONS. *J. Biol. Chem.* **286,** 26050–26060 (2011).

103. Acton, T. B. *et al.* in *Methods in Enzymology* **493,** 21–60 (Elsevier Inc., 2011).

104. Jansson, M. *et al.* High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in Escherichia coli. *J. Biomol. NMR* **7,** 131–41 (1996).

105. Ii, T., Mullen, J. R., Slagle, C. E. & Brill, S. J. Stimulation of in vitro sumoylation by Slx5–Slx8: Evidence for a functional interaction with the SUMO pathway. *DNA Repair (Amst).* **6,** 1679–1691 (2007).

106. Dai, B. *et al.* Tunable assembly of amyloid-forming peptides into nanosheets as a retrovirus carrier. *Proc. Natl. Acad. Sci.* 201416690 (2015). doi:10.1073/pnas.1416690112