RICE UNIVERSITY

# Developing a toolkit for modular adeno-associated virus surface display of peptides and proteins

by

**Nicole Nisha Thadani**

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

**Doctor of Philosophy**

APPROVED, THESIS COMMITTEE

Junghae Suh, Chair
Associate Professor, Associate Chair of Bioengineering
Associate Professor of Biosciences

Jonathan Silberg
Professor of Biosciences

Gang Bao
Professor of Bioengineering
Professor of Chemistry

HOUSTON, TEXAS
May 2019

**Abstract**

# Developing a toolkit for modular adeno-associated virus surface display of peptides and proteins

by

## Nicole Nisha Thadani

Engineering biocomputation in nanotherapeutics is a growing field for the creation of devices that respond to their environment to diagnose diseases or deliver targeted treatments. Viruses are genetically encoded nanoplatforms that come prepackaged with sense-response behaviors allowing them to navigate cellular entry, genome delivery and replication. The field of synthetic virology seeks to enhance viral performance for delivery applications by refactoring viruses into well-characterized domains that can be exchanged or augmented with exogenous functional motifs. Adeno-associated virus (AAV) is a strong candidate for modification through synthetic virology — this vector is relatively well-characterized and has a wide range of potential applications in safe and efficient gene therapy. We sought to develop modular, standardized platforms for the integration of exogenous proteins into the AAV capsid so that biological 'parts' identified in other systems can be translated to enhance AAV-based therapies. By applying protein engineering techniques to study and modify AAV's innate biocomputation, we have developed a series of components that can alter the viral response to external stimulus and expand this platform's capacity for protein and peptide outputs in addition to gene therapy. To address the challenge of genetically modifying the multifunctional virus capsid while preserving viral assembly and transduction, we have identified computational models that may be able to predict vector formation and function from the modified capsid sequence. These models can potentially accelerate the design process for engineered viral nanoparticles through *in silico* screening to remove non-functional variants. This toolkit will facilitate the incorporation of a wide range of proteins in AAV, expanding the vector's capacity for detecting stimuli and responding with a range of diagnostic and therapeutic outputs.

# Acknowledgements

Junghae Suh: Your guidance and unwavering optimism have fueled my scientific growth over the past five years. Thank you for demonstrating a compassionate approach to scientific leadership and for supporting my wide-ranging interests.

My thesis committee: Thank you for encouraging me to think creatively, consider the fundamental questions underlying my research, and tie my work to its potential applications.

My collaborators: I'm grateful for the opportunity to have learned from your expertise. Thank you to the members of Peter Wolynes's group for teaching me about coarse-grained modeling and guiding my frustratometer analyses. Thank you to Faruck Morcos and Qin Zhao for your many thoughtful conversations and coevolution analyses of AAV. Thank you to Ciaran Lee and Buhle Moyo for your Cas9 engineering on the leucine zipper project and for cheerfully troubleshooting the many roadblocks we encountered.

Annicka: Thank you for always seeing my strengths and helping me look on the bright side - your friendship and support has been invaluable throughout this graduate school adventure.

Joanna, Kiara, and Jiayu: Your dedication and enthusiasm inspire me. Thank you Joanna for your many contributions to the leucine zipper project and for deep conversations (and constant support) in lab. Thank you Kiara for your time on my modeling projects and for showing me the weirdest, most entertaining videos. Thank you Jiayu for bringing your energy and funny stories every day. Y'all make research infinitely more enjoyable.

Suh Labbers: Thank you for the scientific (and definitely not scientific) insights, commiseration about research frustrations, dad jokes, lab lunches, office arts and crafts, and the many snacks.

Ian, Charles, Santi, Mary, Eric, Gisele, Dessy, and Jason: Thank you for adventuring and experiencing Houston with me these past five years. Glad I got to share grad school with y'all.

Valhalla Bartenders: Thank you for making the bar an incredible place and an essential community.

My parents, Heeru and Anita: Thank you for your constant encouragement and belief in me. I am forever grateful for the opportunity to pursue my dreams and know that you always have my back.

Matt: Thank you (and Lincoln) for being my number one fan and for dragging out the fun parts of me from under the piles of work. Couldn't have done it without you.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Adeno-associated virus: A genetically-encoded nanoplatform

Adeno-associated virus (AAV) is a small non-pathogenic parvovirus that infects a wide range of mammalian cells. AAV is widely considered a promising candidate for gene therapy — AAV-based therapies recently gained regulatory approval in the U.S. and in Europe [1–3]. AAV's prominence as a therapeutic is partially due to its benign safety profile - the vector has mild immunogenicity as virus replication requires coinfection with a helper virus (adenovirus or herpes simplex virus) and infection is non-pathogenic [4]. AAV's small size, structural simplicity and value as a therapeutic vehicle make it an ideal candidate for studying innate viral sensing-response behaviors and engineering these behaviors to change the inputs and outputs of viral infection.

### 1.1.1 AAV structure

AAV is composed of a non-enveloped capsid packaging a single-stranded DNA genome flanked by two inverted terminal repeats (ITRs). The viral genome contains two genes: *rep*, which encodes four overlapping non-structural proteins required for virus replication and *cap*, which encodes the three structural capsid proteins and the Assembly-Activating Protein (AAP) in an alternate reading frame (Figure 1.1). The AAV genes are governed by the promoters p5, p19, and p40. In gene delivery vectors, this genome is replaced by the desired cargo flanked by ITRs, and *rep* and *cap* are delivered in trans for assembly [5].

The AAV genome expresses the *rep* gene to produce Rep78, Rep68, Rep52, and Rep40, and the *cap* gene to produce VP1, VP2, VP3, and AAP. The Rep proteins have roles in gene transcription, capsid assembly, DNA packaging, and genomic integration. The Cap proteins VP1, VP2 and VP3 (molecular weights of 87, 73, and 61 kDa respectively) assemble in a 1:1:10 ratio to form a 60-mer capsid with icosahedral (T=1) symmetry. VP1 and VP2 share the VP3 domain at their C-termini which forms the exterior structure of the capsid. This domain makes up the $\beta$-barrel motif with $\beta$-sheets connected by hypervariable interstrand loops that is typical of parvoviruses. The N-termini of VP1 and VP2 pack inside the capsid and extrude during transduction in response to external stimuli. The unique N-terminus of VP1 (VP1u) contains nuclear localization sequences (NLSs) and a phospolipase A2 (PLA2) domain that facilitates endosomal escape. The role of the unique region of VP2 (VP2u) is unknown [5].

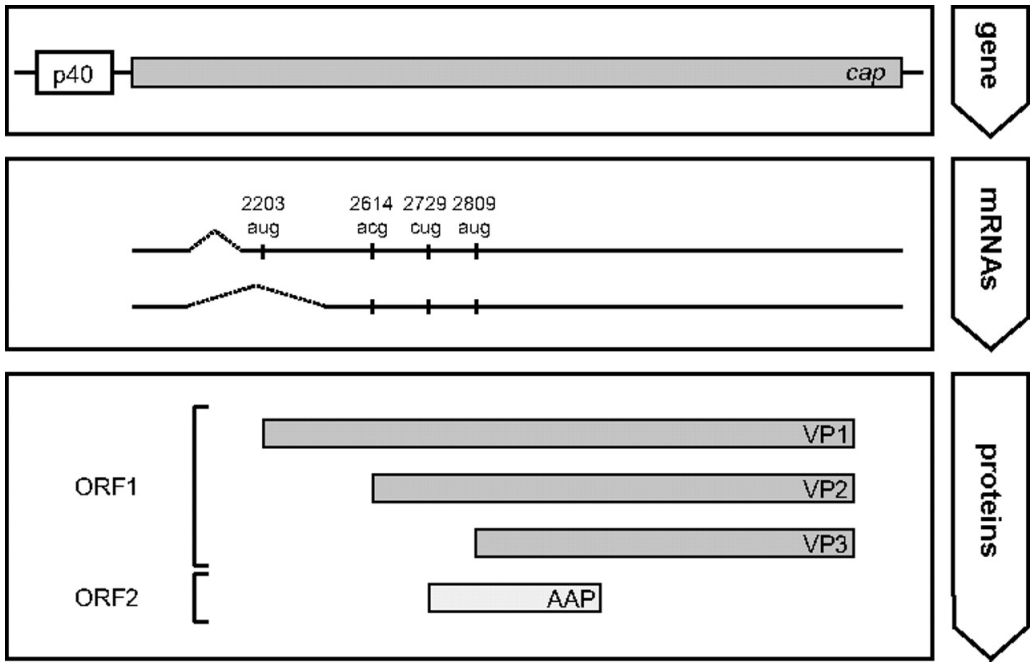



**Figure 1.1:** Organization of VP1, VP2, VP3, and AAP on the AAV2 *cap* gene. Reprinted with permission from [6].

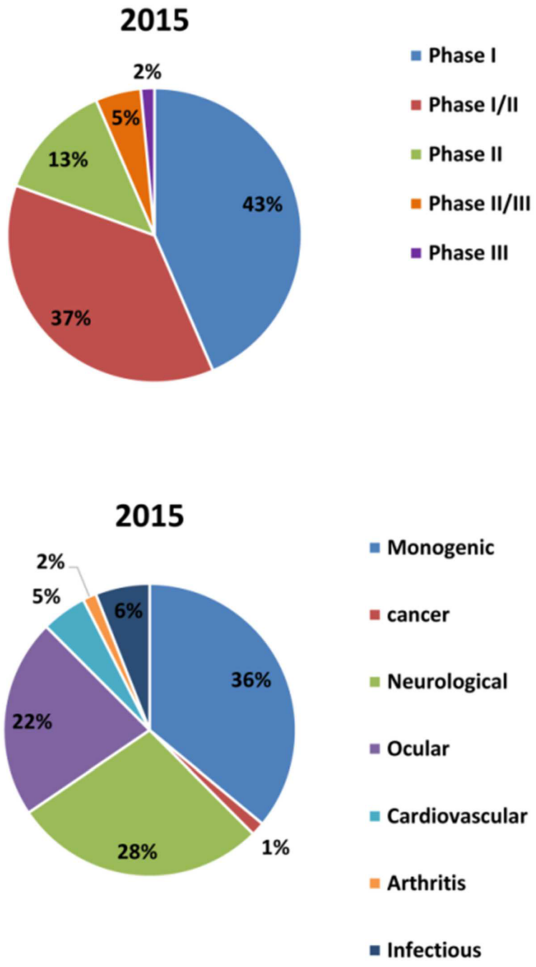



**Figure 1.2:** Statistics showing the clinical phases [top] and disease targets [bottom] of clinical trials performed with AAV gene therapy in 2015. Adapted with permission from [7].

### 1.1.2 AAV in gene therapy

Gene therapy focuses on the manipulation of gene expression to achieve therapeutic outcomes. This is done by delivering genes, or the means for modifying genes and their expression patterns, to disease sites. Gene therapy approaches were introduced as cures for genetic disorders, and their applications have since expanded to encompass a variety of acquired diseases including cancer, heart disease, and diabetes [8]. AAV-packaged therapies have treated these and a variety of other diseases in clinical trials (Figure 1.2).

AAV is commonly used as a gene therapy vector because of its many favorable characteristics. Like other viral vectors, AAV has been evolutionarily optimized to transduce cells efficiently. AAV is approximately 25 nm in diameter, allowing the virus to directly enter cell nuclei and deliver its genome [9]. The AAV genome persists in cell nuclei as episomal DNA, producing sustained gene expression through multiple cycles of cell division [10]. Twelve different serotypes of human AAV have been identified thus far, each with unique tissue tropisms [11, 12].

While AAV has been used in a broad spectrum of gene-delivery therapeutics, the virus has some flaws as a vector. The capsid is susceptible to immune recognition. A large percentage of the human population has been previously exposed to the virus, resulting in detectable neutralizing antibodies in their blood [13]. In addition, the different AAV serotypes each act on several types of tissue, so specificity of targeting is limited [12]. When AAV transduces cells, the virus is further challenged by endosomal and proteosomal degradation machinery [14]. Finally, the virus can only package 4.7 kb of DNA, so therapeutic cargo is heavily restricted in size [5].

Many investigators have taken up the challenge of improving AAV as a therapeutic gene delivery vector. Typically, engineering of viral vectors is conducted at the genetic level, although chemical modification is also used to functionalize capsid surfaces [15]. Researchers have focused on engineering the AAV capsid to reduce immunogenicity, improve intracellular trafficking to the nucleus, and optimize AAV specificity. Rational design approaches have incorporated peptide ligand sequences to alter receptor targeting, removed tyrosine residues to mitigate proteosomal degradation, and mutated antibody-binding epitopes to reduce serum neutralization [14]. Directed evolution approaches have also been used to identify improved AAV vectors from randomly mutated libraries of AAV capsids [16–19]. These libraries are processed through selections to screen for antibody evasion and improved tissue tropism [14]. These studies have produced some promising candidates for gene therapy, but much is still unknown about the impact of mutating different AAV capsid domains on the virus's ability to transduce cells and deliver genes. Understanding AAV's innate capacity to respond to its environment during infection may provide insight into strategies for engineering biocomputation into AAV.

## 1.2 AAV life-cycle dynamics

The AAV capsid undergoes a series of structural changes as it assembles, binds to cell receptors, travels through the cytoplasm, and enters the nucleus. These changes allow the virus to respond to its environment, evading biological barriers to nucleic acid delivery in the extracellular space and cytoplasm and releasing the genome at its destination. Such transitions are possible through the multifunctional nature of the capsid proteins. This section describes each stage of the virus life cycle with a focus on the cascade of sensing-response behaviors required to trigger successful viral formation and infection. Data referenced is for AAV2 unless otherwise stated.

### 1.2.1 Viral assembly

AAV capsid assembly and genome packing occur in the cell nucleus, and require the Rep and Cap proteins, AAP, Ad proteins, and cellular co-factors. First, the *rep* and *cap* genes are transcribed in the nucleus with the help of adenoviral transcription factors, then spliced and translated in the cytosol. After the Rep and Cap proteins are synthesized they traffic to the nucleus (although a significant fraction of VP3 remains in the cytosol) [20]. Capsid proteins likely traffic to the nucleus in trimer assemblies, as these assemblies are found in the cytoplasm [21] and trimers are the most stable intermediate assembly in a variety of parvoviruses [22].

#### Capsid protein assembly

Cap proteins are capable of aggregating in solution and in the cytoplasm, but require nuclear localization to assemble into capsids, suggesting a dependence on nuclear cofactors [21, 23]. In the nucleus, Cap proteins bind to the nuclear signalling proteins Nucleolin and B23/Nucleophosmin and enter nucleoli [24, 25]. The first indicator of AAV2 capsid assemblies, as detected by antibody A20, appear in the nucleolus [20]. One of four identified basic regions on the capsid proteins (BR4) is essential for this early stage of assembly. While the basic regions are believed to play a role in nuclear trafficking, AAV capsids with a mutated BR4 region contain a similar amount of the VP proteins in the nucleus as compared to wild-type, indicating that this region may instead play a role in intersubunit or essential cofactor interactions [26].

#### Assembly-activating protein

The assembly-activating protein (AAP) is a 23 kDa protein translated from an alternate reading frame and a nonconventional start codon on the *cap* gene. AAP is found in the nucleolus along with the Cap proteins during virus assembly. AAP contains several redundant nuclear localization signals (NLSs) and nucleolar localization signals (NoLSs) that allow the protein to translocate to the nucleus. When these signals are ablated, viral capsid formation is heavily impaired [27]. For AAV2, when AAP expression is ablated, VP3 does not traffic to the nucleolus. Even when nucleolar localization of VP3 is recovered using a nucleolar localization signal tag, capsid formation does not occur, indicating that AAP plays a role both in nucleolar localization of VP3 and in donwstream assembly [6]. AAP has also been demonstrated to promote VP stability and to increase VP-VP interactions in co-immunoprecipitation assays [28].

Some specific interaction domains required for AAP function have been identified. AAP contains two hydrophobic domains at its N-terminus that are necessary for assembly. Mutations at these regions reduce AAP interactions with the capsid proteins. In addition, mutations at the C-terminus of VP3 reduce AAP interactions with the capsid proteins and ablate assembly, suggesting that the C-terminus of VP3 interacts with AAP. This conclusion is reinforced by evidence of competition between the antibody B1 (which binds the C-terminus of the capsid proteins) and AAP. Antibody studies show that AAP binding promotes capsid protein conformations indicative of assembly, suggesting that AAP may serve as a scaffold [29].

Many AAV serotypes require AAP for virus production (AAV1, 2, 8, and 9 have been tested) [30]. AAP is not required for AAV4, 5, and 11 assembly, but promotes capsid assembly resulting in 10-fold higher titers [31]. AAPs are somewhat interchangable between virus serotypes — the AAV2 AAP stimulates virus production for AAV1, AAV8 and AAV9 but not for AAV5 (Figure 1.3). Based on studies of AAP-dependence in the AAV phylogeny, it appears that AAP-dependent serotypes may have fewer evolutionary constraints on the residues at VP-VP interfaces, as AAP facilitates VP-VP binding [28].

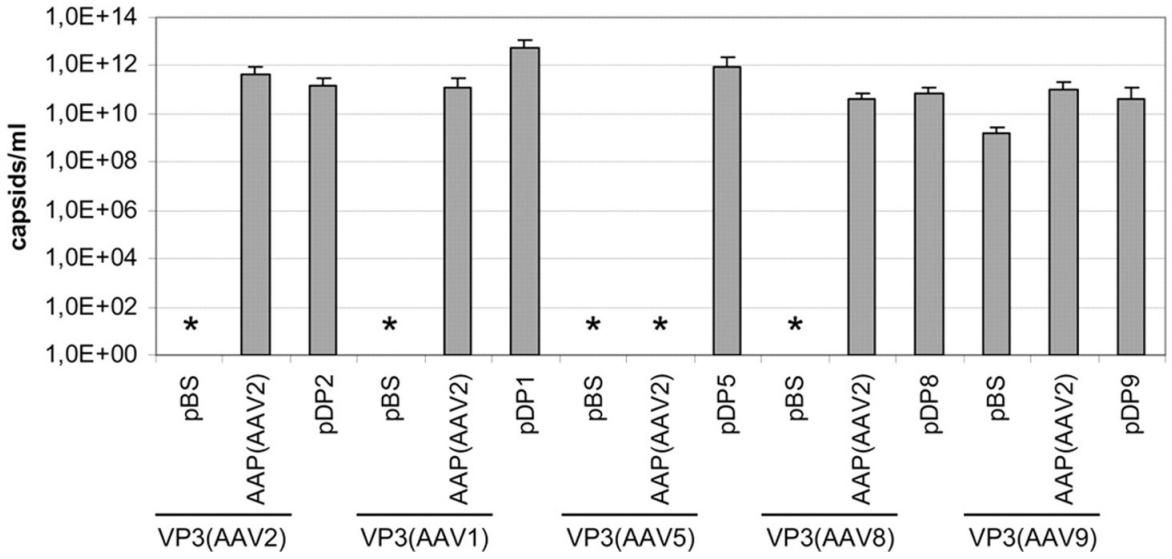



**Figure 1.3:** HEK293t cells were transfected with plasmids expressing VP3 of various serotypes and either an empty plasmid (pBS), a plasmid expressing AAP2, or a plasmid expressing the full *cap* gene of the indicated serotype (pDP). Capsid titers were quantified by ELISA. AAV9 appears to form without AAP, but further analysis showed that the VP3 sequence of AAV9 contains a start codon that expresses a truncated AAP, and when this is ablated AAV9 does not form. Reprinted with permission from [30].

### Genome preparation

The AAV genome is replicated through a single-stranded displacement mechanism. First, reverse strand synthesis occurs, creating a covalently attached daughter strand. Rep proteins then cleave the parental strand before the ITR to resolve the end of the daughter strand and reinitiate transcription (Figure 1.4). Rep78 and Rep68 exhibit sequence specific endonuclease and helicase activity to nick the parental strand [32, 33]. These large Rep proteins also interface with cellular, adenoviral, and HSV ssDNA binding proteins, which play a role in genome replication [34].

### Genome packaging

Genome packaging is an essential step in viral maturation, and is often incomplete with 50%+ of capsids remaining empty [36]. After Cap proteins assemble, they colocalize with all four Rep proteins, the AAV genome, and cellular cofactors in the nucleoplasm [20,37]. Genome packing occurs in these replication centers, which colocalize with adenovirus replication centers and co-opt their machinery in coinfected cells [38]. The DNA genome associates with immature formed capsids rapidly, but full encapsulation and protection of the genome occurs over several hours [39]. The Rep52 and Rep40 proteins are essential in genome packaging, and act as helicases on the DNA genome. DNA has been shown to pack into capsids from the 3' end and the Rep52 and 40 helicases act in the 3'→5' direction, suggesting that the Rep52 and 40 helicases play a role in inserting the DNA genome into the capsid [40]. These small proteins interact with the large Rep proteins, which stay bound to DNA as it is packaged. Rep proteins interact with each other and with the VP proteins even in the absence of viral genomes, so it is likely that the Rep proteins form a complex on the outside of the viral capsid that facilitates genome packaging [41, 42].

Mutational studies have been used to better understand the genome entry point into the capsid. Mutations at the 5-fold axis pore have ablated genome packing and Rep protein binding, so insertion may occur

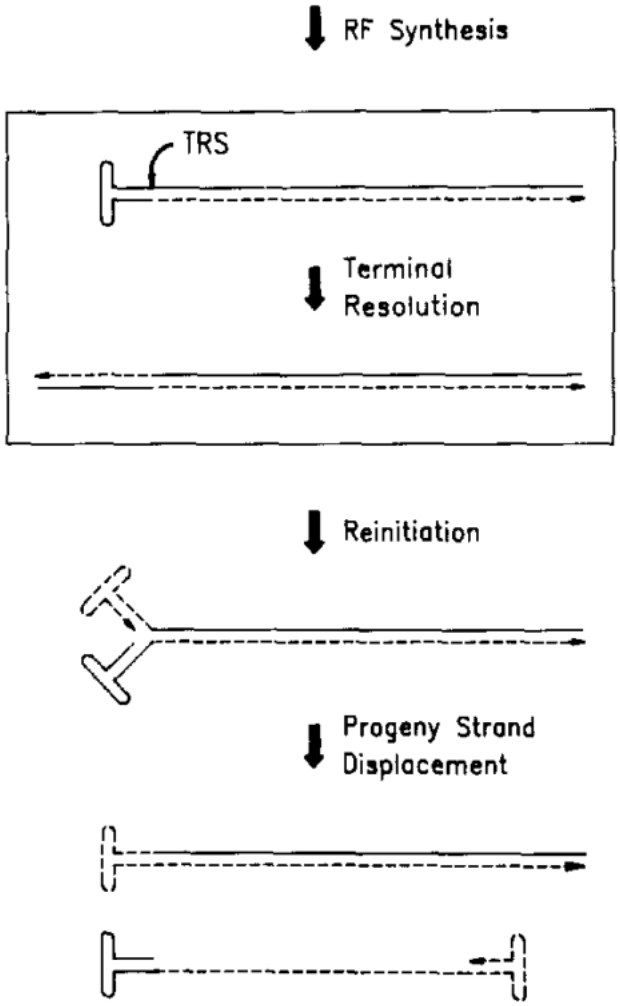



**Figure 1.4:** For the AAV genome to replicate, the 3' ITR acts as a primer for elogation. Then, the large Rep proteins facilitate nicking of the parent strand (boxed). Next the parent strand acts as a primer to replicate the 3' ITR, elongating the daughter strand. Finally the strands separate. Reprinted with permission from [35].

at this location [43, 44]. The sequence of the HI surface loop (which overlaps with the neighboring VP at the 5-fold interface) is also a factor in genome packaging [45]. Mutations in a variety of other capsid regions appear to limit genome packing, particularly those at the base of the three-fold protrusions [46,47]. One such mutation, R432A, has an impact despite being structurally buried, likely due to its impact on overall capsid stability and other neighboring residues [48]. Genome length also has a strong impact on AAV packaging, with larger genomes over the 4.7kb packaging limit resulting in more empty capsids [49].

### 1.2.2 AAV extracellular immune interactions

The immune system challenges AAV on several fronts as the mature capsid travels through the body. The capsid primarily interacts with the immune system through antibody binding. Immunoglobulin G antibodies confer pre-existing immunity to wild-type AAV in a majority of the human population. However, this immunity varies greatly between virus serotypes, which are distinguished by their lack of antibody cross-

reactivity (Figure 1.5). AAV2's antibody interaction profile is the best known. Antibodies have been identified that interact with the VP1, VP2, and VP3 protein domains and that interact with domains created by subunit assemblies. Epitopes of these antibodies have been mapped onto the capsid surface, allowing for the generation of capsid mutants that evade antibody detection [50, 51]. Antibody neutralization occurs at the cell binding and intracellular trafficking points of transduction [52].

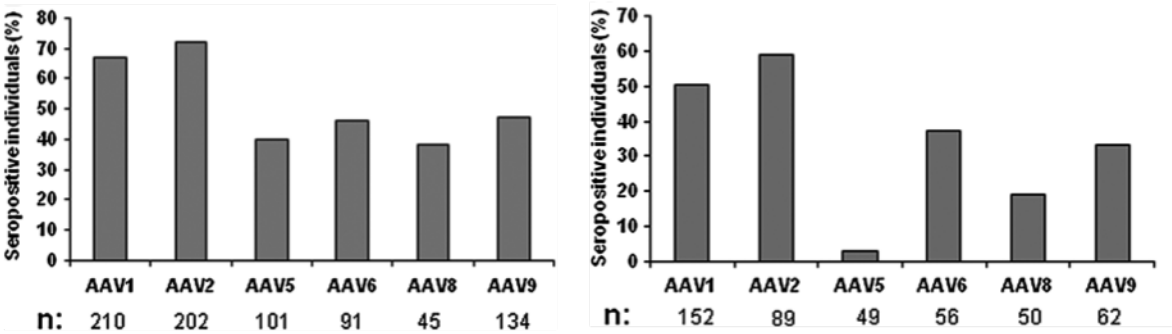



**Figure 1.5:** Percentage of sera samples with IgG seroprevalance (left) and percentage of sera samples with neutralizing antibodies (right) for the indicated AAV serotypes. Sample numbers below. Adapted with permission from [13].

### 1.2.3 Cell receptor binding

One reason for AAV's popularity as a gene delivery vector is the ability of the different serotypes to target a wide range of cellular receptors for cell entry (Table 1.1). Cell receptors are the initial contact for AAV entry into cells through clathrin-mediated endocytosis. Heparan sulfate proteoglycan (HSPG), sialic acid, and galactose have been identified as significant receptors for different serotypes [53–55]. In addition, a recently discovered receptor, AAVR, was shown to be essential for AAV infection in all the tested serotypes [56]. The HSPG interaction with AAV2 has been heavily characterized, and the binding domain on the capsid has been identified at the peaks surrounding the three-fold axis of symmetry [57]. As AAV2 binds HSPG, it undergoes a conformational shift, with the 2-fold axes and 5-fold pores changing shape to prepare for downstream steps in transduction and genome delivery [57]. Sialic acid (AAV1 and 6) and galactose (AAV9) binding sites have also been identified in a pocket at the base of the peaks surrouding the three-fold axis [58, 59]. These receptors in conjunction with co-receptors mediate capsid endocytosis. In AAV2, cellular entry requires the activation of the Rac1 protein that acts with the phosphatidylinositol-3 (PI3) kinase pathway to activate viral transport along the cytoskeleton [60].

**Table 1.1:** Primary Cell Receptors for AAV Serotypes. Adapted from [5].

| **Serotype** | **Receptors/co-receptors** |
|---|---|
| *AAV1* | AAVR, $\alpha 2-3/\alpha 2-6$ N linked SA |
| *AAV2* | AAVR, HSPG, FGFR1, HGFR, integrins, 37/67 kDa LamR |
| *AAV3* | AAVR, HSPG, 37/67 kDa LamR |
| *AAV4* | $\alpha 2-3$ O linked SA |
| *AAV5* | AAVR, $\alpha 2-3$ N linked SA, PDGFR |
| *AAV6* | AAVR, HSPG, $\alpha 2-3/\alpha 2-6$ N linked SA |
| *AAV7* | Not determined |
| *AAV8* | AAVR, 37/67 kDa LamR |
| *AAV9* | AAVR, Galactose, 37/67 kDa LamR |

### 1.2.4 Intracellular trafficking

Once AAV enters a cell, the capsid must travel to the nucleus while protecting its genome. Nuclear entry in particular is a rate-limiting step for the AAV vector, leading to slow expression and vastly impaired transduction in many cell types [61,62]. In addition, impaired trafficking can lead to proteosomal degradation of the capsid and cellular antigen presentation, targeting transduced cells for destruction [63].

#### Endosomal trafficking

The first stage of AAV transport occurs in the endosome. AAV2 has been found in both the early and late-stage endosome, and the timing of endosomal escape varies by cell type and multiplicity of infection (MOI) [64]. While AAV2 can escape early endosomes, AAV8 appears to stay in the endosome longer and interact with proteosomal degradation machinery in its infection pathway [65]. The microtubule network transports endosomal capsids towards the nucleus [66]. The capsid requires acidification of the endosome and digestion by endosomal cathepsins for successful transduction [67, 68]. This endosomal environment induces surface extrusion of the N-terminal domains of VP1 and VP2, a conformational change that has been induced in the lab by heat [9,69]. Exposure of the N-terminal domain of VP1 is necessary for transduction, likely due to its PLA2 domain for endosomal escape and NLS regions for nuclear transport [70]. This region of VP1 also contains endosomal sorting and signal transduction domains that are likely required [71]. The exact path of N-terminal extrusion is unknown, but mutational analysis and structure modeling have illuminated parts of the mechanism. Mutation of residues at the 5-fold pore leads to reduced N-terminal extrusion, suggesting that this portion of the virus structure plays a role in N-terminal dynamics, although some of these mutations also have broad impacts on capsid structure [43, 70]. Mutation of the HI surface loop also impacts N-terminal extrusion [45, 57]. Prevention of conformational change at the 2-fold capsid interface also limits n-terminal extrusion, so this domain may also play an essential role in N-terminal dynamics [72]. Structural modeling of the AAV1 capsid led to a hypothesized model of N-terminal extrusion in the late endosome. First the PLA2 domain unfolds so that the N-terminus of VP1 can extrude from the five-fold pore. Then the PLA2 domain partially refolds to digest the endosome membrane and release the capsid

(Figure 1.6).

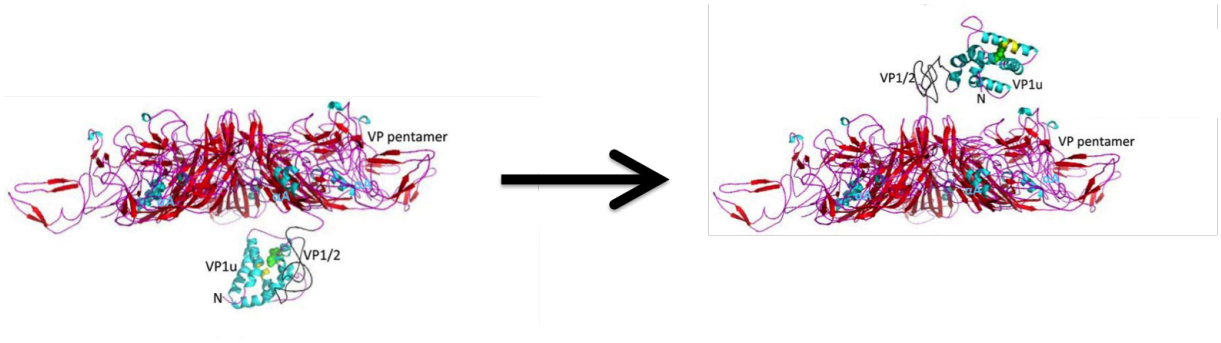



**Figure 1.6:** Structural modeling results indicate that when the AAV1 capsid enters the late endosome, the N-terminus of VP1 relaxes its secondary structure and extrudes from the capsid. Then the PLA domain partially refolds to regain enzymatic activity. Adapted with permission from [73].

### Cytoplasmic trafficking

After AAV leaves the endosome, the microtubule network of the cell facilitates unidirectional AAV transport towards the nucleus, although some capsids appear to move independently through brownian motion [66, 74]. Interactions between AAV capsids and dynein, other microtubule machinery, and protein chaperones have been observed, but the necessity of these interactions for succesful transduction is unclear [75–77]. In fact, the microtubule transport pathway may trap AAV in the perinuclear space, limiting successful nuclear entry [78]. The VP1 N-terminal region seems to play a role in cytoplasmic trafficking, as indicated by mutation studies [71]. In the cytoplasm, the viral caspid is subject to proteosomal digestion [79,80]. Inhibition of proteosome activity increases AAV transgene expression [81]. Proteosomal activity on the capsid was found to be mediated by ubiqutination signaling through the activity of EGFR protein tyrosine kinase on capsid tyrosines [82, 83]. This activity is reduced by mutation of tyrosine residues on the capsid surface, enhancing transduction [72].

### Nuclear entry

The few AAV serotypes characterized for nuclear entry vary in entry rates and efficiency, suggesting different mechanisms of traversing the nuclear membrane. AAV2 is dependent on a functional nuclear pore complex and interacts with importin for cellular entry [84, 85]. In this serotype, one of four identified basic regions on the viral capsid sequence, BR3 (located on the N-terminus of VP2) is essential for nuclear entry. When BR3 is knocked down function can be recovered by introduction of a NLS [26]. AAV2 can transduce nuclei of a variety of cell types, even outside the cellular environment [86]. Nuclear entry is enhanced by the presence of adenovirus capsids [87]. AAV8 appears to enter the nucleus at a different rate than AAV2, potentially due to a different entry mechanism [88]. Some parvoviruses enter nuclei by physically damaging the nuclear envelope, so it is possible that some serotypes of AAV also create a non-nuclear pore complex based entryway [89].

## 1.2.5 Genome delivery

The final step in the AAV life cycle is unpackaging the viral genome and preparing the gene for expression. This process is particularly slow for AAV, resulting in a lag time of 24-48 hours for initial gene expression and

persistence of intact capsids in the nuclei for weeks [61, 90–92]. AAV relies on cellular proteins for assistance with uncoating and DNA processing to allow transcription.

### Capsid disassembly

AAV capsids appear to fully disassemble in the nucleus to release the genome, as indicated by nuclear antibody studies [9]. However, this process is highly inefficient, and intact capsids persist in the nucleoli for up to several years [93–95]. *In vitro* studies simulating denaturation with heat and *in vivo* studies indicate that capsid denaturation occurs at the same time as genome release [96]. Denaturation occurs in the nucleoplasm, as nucleolar entry blocks this change [95]. This transformation is potentially restricted by phosphorylation by CDK2/CyclinA kinase, which binds to the VP1 N-terminus at BR3. This kinase shows reduced activity in the presence of AAV, suggesting evolved inhibition to enhance transduction [96]. The degree of capsid disassembly in the nucleus likely varies across serotypes, as different serotypes exhibit diverse stability characteristics [97]. In addition, capsid disassembly *in vitro* is impacted by the length and binding state of the packaged DNA [98].

### Genome processing

Once the AAV capsid releases its genome, the genome must undergo second-strand synthesis before transcription can occur. Second strand synthesis is the primary method of gene transcription, but it is possible that strand annealing between opposite-sense strands plays a role in recombinant AAV gene delivery [99, 100]. In wild-type AAV infections, this process is facilitated by adenovirus protein expression [101, 102]. Adenoviral proteins dephosphorylate a cellular protein, FKBP52, which binds to the 3' ITR on the genome and blocks second-strand synthesis [103, 104]. In recombinant AAV infections, repressing the phosporylation of FKBP52 increases transduction [104]. Adenoviral proteins also repress the cellular Mre11-Rad50-Nbs1 (MRN) repair complex, which targets the ITRs as AAV genome foci colocalize with repair proteins [105–107]. Some evidence suggests that the AAV capsid itself also plays a role in second strand synthesis after disassembly. Mutants of certain capsid residues result in a defect in transduction that is rescued by delivery of a double stranded genome, although genome uncoating is equivalent to wild type [108]. The virus capsid may also play a role in downstream transcription of the genome, as mutants near the two-fold interface affect transcription levels independent of second strand synthesis [109]. After second strand synthesis, viral genomes co-opt cellular machinery to self-circularize, as dsDNA free ends are unstable in the nucleus [110, 111].

## 1.3 Biocomputing nanoparticles*

Nanoscale platforms designed as drug and gene delivery vectors or diagnostic agents have the potential to dramatically improve the quality of life of patients suffering from a wide range of debilitating diseases, such as cancer, cystic fibrosis, and Alzheimer's Disease [112]. Delivering drugs, genes, or imaging agents both effectively and specifically to desired target tissues and cells could critically improve biological and clinical understanding of a pathological condition and provide a means for halting or even reversing the disease. Nanoplatforms have also demonstrated utility as detection agents, precisely identifying pathogens and other disease markers, environmental toxins, and hazardous materials. As methods have improved to synthesize a vast array of nanoplatforms with more controllable physicochemical properties, investigators have been

*This section has been adapted from Evans, A. C. and Thadani, N. N., *et al.* (2016). Biocomputing Nanoparticles for Therapeutics and Diagnostics. *Journal of Controlled Release*.

pushing the boundaries of what sophisticated functions can be performed by the nanomaterials. In particular, a new breed of nanotechnologies able to perform "biocomputation" is being actively developed. For the purposes of this review, biocomputation is defined as the act of sensing and, importantly, computing biomolecular inputs either through algorithmic trees or continuous functions, and in response generating functional outputs. The concept of biocomputation, with roots in electrical engineering, has been applied with much success in the field of synthetic biology. Conceptualizing and designing genetic circuits as platforms that compute biomolecular inputs through various methods, such as Boolean logic, have been explored extensively [113–115]. Applying the abstractions of computation to the design of nanoscale platforms, either biologically derived or synthetic, is a relatively new approach to thinking about how these functional nanomaterials can be engineered to improve and expand their capabilities.

Promisingly, a plethora of inputs and outputs are compatible with nanomaterial-based biocomputation (Figure 1.7). The biomolecular inputs can be small molecules, enzymes, or nucleic acids. In addition to these intrinsic inputs, a variety of inputs exogenously applied to a biological system, such as light or magnetic fields, can be accepted as stimuli. The inputs can be computed via a number of different logic operators, such as AND, OR, and NOT gates. In response to detecting and computing inputs, the nanoplatforms carry out specific functions, such as delivering transgenes into cells or producing a fluorescence signal. Thus, by enabling nanomaterials to conduct defined computations, platforms can be designed to carry out desired complex tasks, such as delivering desired doses of drugs only to sites of disease characterized by combinations of target biomarkers. Then by characterizing the engineered platforms quantitatively and modeling their behavior mathematically, the engineered biocomputing nanomaterials can become not only controllable but also functionally predictable. This potential to predict the output of a nanoplatform given a set of biomolecular inputs is what makes the biocomputational approach impactful to the future of drug delivery and bioimaging. Much of the work on stimulus-responsive nanomaterials can be easily re-conceptualized through the lens of biocomputation; it would require categorizing the type of computation being conducted by the nanomaterial and quantitatively characterizing the input-output curve under various conditions.

This review will cover some of the recent advances in designing and testing nanoplatforms that conduct biocomputation. The definition of a biocomputing nanoplatform can encompass any nanomaterial or nanoparticle that detects and processes an input signal and outputs a defined response. This review will focus on the more complex biocomputing nanoplatforms developed to date that explicitly incorporate logic gates and often require more than one input for activation. Many different types of nanomaterials are being programmed as biocomputing platforms for a range of applications (Table 1.2). To date, the main challenges being addressed by biocomputing nanoplatforms are targeted and controlled delivery of drugs or genes, detection of analytes using *in vitro* diagnostics, and *in vivo* molecular imaging. The biocomputing approach promises to enable more precise, more selective, and potentially more quantitatively predictable activation of the desired functional output (e.g. delivery of a therapeutic payload or release of a fluorescence signal). The variety of intrinsic and extrinsic input stimuli, output responses, and computational methods used by the different platforms will be discussed.

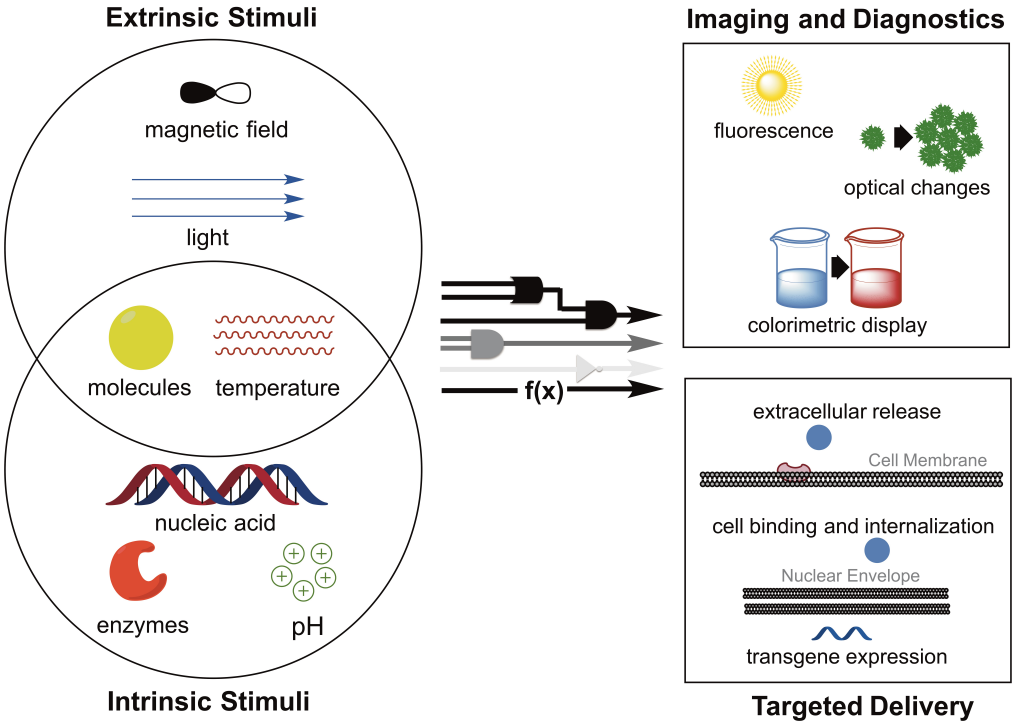



**Figure 1.7:** Biocomputing nanoplatforms accept inputs that are either intrinsic or extrinsic to biological systems, compute the inputs using defined functions, and produce outputs that manipulate the biological environment (therapeutics) or act as an externally detectable readout (diagnostics).

**Table 1.2:** Applications of biocomputing nanoplatforms

| Nanomaterial Platform | Applications |
|---|---|
| *Polymeric* | Inflammatory disease targeting, cancer treatment, targeted imaging |
| *Gold-coated* | pH and heavy metal sensing, nucleic acid sensing, "lab-on-a-nanoparticle" |
| *Nucleic acid-based* | Fluorescent and colorimetric biosensing, antibody delivery |
| *Enzyme-based* | Detection of biochemical substrates |
| *Magnetic* | Magnetic resonance imaging, cell sorting |
| *Viral* | Cancer treatment, toxin sensing, cell targeting |

### 1.3.1 Operators used by biocomputing nanoplatforms

In the most ideal manifestation, nanomaterial-based computation translates inputs into outputs in a quantitatively predictable manner. Researchers have primarily focused on developing systems with binary outputs, similar to the logic gates found in silicon-based computing. These logic gates have been particularly advantageous when applied to biological systems for building complex protein-based signal transduction networks with sophisticated computing-like behavior [116]. These binary systems must be able to detect inputs above a certain threshold level and reject those below the threshold. Many biocomputational nanomaterials have been programmed as AND gates that require detection of two different stimuli to activate. These AND gate nanomaterials can either integrate simultaneous stimuli, or they can utilize consecutive stimuli to activate in phases as they transport to their biological targets. Other biocomputational nanomaterials function as OR gates, NOT gates, or combinations of multiple gates. Some biocomputational nanomaterials behave as tunable devices, where the level of output is determined by the level of input detected. In other words, tunable nanodevices use a continuous mathematical function to connect the level of input signals to the precise level of output, allowing for calibrated responses. The various biocomputing nanoplatforms have different strategies for translating multiple stimuli into a single output, including system assembly, disassembly, and transformation (Figure 1.8). The remainder of the review will cover examples of these three different strategies.

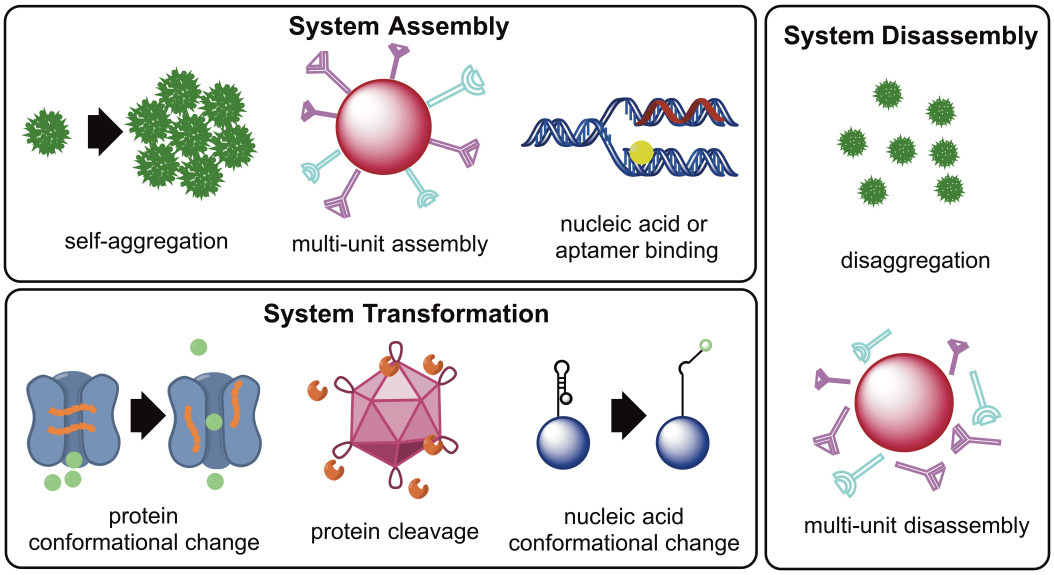



**Figure 1.8:** Methods of nanoplatform information processing fall into three broad categories: system assembly, system disassembly, and system transformation.

### 1.3.2 System assembly as means of integrating inputs

Many biocomputational nanomaterials integrate single or multiple inputs through the assembly of individual components, a process termed "system assembly". System assembly of nanoparticles has been applied in several pioneering studies to monitor biomolecular interactions, as the aggregation of small metallic nanoparticles can influence their optical and magnetic properties [117–121]. Nanoparticles tethered to DNA, proteins, and enzymes can thus be used to optically observe proximity. In this review, we will focus on examples of system assembly where the output is produced based on a logical computation from a set of biomolecular

inputs.

Oftentimes, system assembly can be mediated by DNA hybridization. For example, one logic gate-based strategy with DNA nanostructures exploits the hybridization of two different input segments of single-stranded DNA, thus operating as an AND gate, to create a longer near-complementary double-stranded DNA fragment. The combined strand then uses toehold-displacement to assemble onto a gold nanoparticle functionalized with DNA fragments. This unique assembly process displaces a fluorophore-conjugated DNA strand, ultimately resulting in the generation of fluorescence as the output [122]. The platform may be used to detect multiple DNA inputs for achieving DNA-based computation. In another report, similar gold nanoparticles functionalized with single-stranded DNA have been developed to function as target DNA sensors [123]. The detection of target single-stranded DNA facilitates assembly of the nanoparticles into multi-nanoparticle aggregates, a reaction that can be visualized through a solution color change. The platform can be designed to require detection of two different target DNA strands (AND gate), or the detection of one of two DNA segments (OR gate) to signal the presence of pathogen DNA in a sample.

Similarly, DNA aptamers have been attached to the surface of nanoparticles so that aptamer conformational shifts triggered by binding of small molecules can be used to induce the hybridization of near-complementary DNA strands and trigger nanostructure aggregation. For example, one AND gate nanoplatform designed for small molecule sensing used aptamer-modified gold nanoparticles to detect both adenosine and cocaine, inducing nanoparticle aggregation that resulted in a solution color change as the output [124]. Overall, DNA-based aggregation systems utilize the temperature stability and well-characterized hybridization behavior of nucleic acids to create precise logic gate-based nanodevices.

System assembly-based biocomputing nanoplatforms have also been generated without the help of DNA mediators. One nanoplatform relies on spiropyrans to function as an AND-gate, detecting UV light and metal ions as dual inputs [125]. In the presence of the light stimulus, spiropyrans conjugated to the surface of gold nanoparticles can bind metal ions in a multivalent fashion, thus inducing aggregation of the gold nanoparticles and creating a solution color change output. This system can be used to quickly detect a variety of metal ions in solution with a visually detectable change. Another approach utilizes pH and temperature-responsive polymers on the surface of magnetic nanoparticles to relate these stimuli to a continuous range of aggregation outputs [126]. This platform allows cells which contain aggregated nanoparticles to be separated from cells that do not contain aggregated nanoparticles, since only aggregated nanoparticles can be separated with a magnetic field. Other approaches use protease cleavage domains to detect matrix metalloproteinases (MMPs), a class of enzymes elevated in various diseases [127, 128]. One platform uses MMP-cleavable peptides on the surface of iron oxide nanoparticles (Figure 1.9) [129]. These peptides act as AND or OR gates, cleaving in the presence of target MMPs and inducing nanoparticle aggregation to change the magnetic properties of the particles. Such MMP-sensing nanoparticles can be used as *in vivo* diagnostic agents using magnetic resonance imaging.

In summary, several biocomputing nanoplatforms use assembly-based strategies, such as DNA hybridization or multi-nanoparticle aggregation mediated by polymers, peptides or DNA. These assembly processes can be triggered by a wide range of stimuli, such as small molecules, enzymes, UV light, temperature and pH. As nanoparticle aggregates often have different fluorescence or absorbance properties as compared to their individual components, the system assembly approach is primarily used to generate diagnostics and other detection tools.

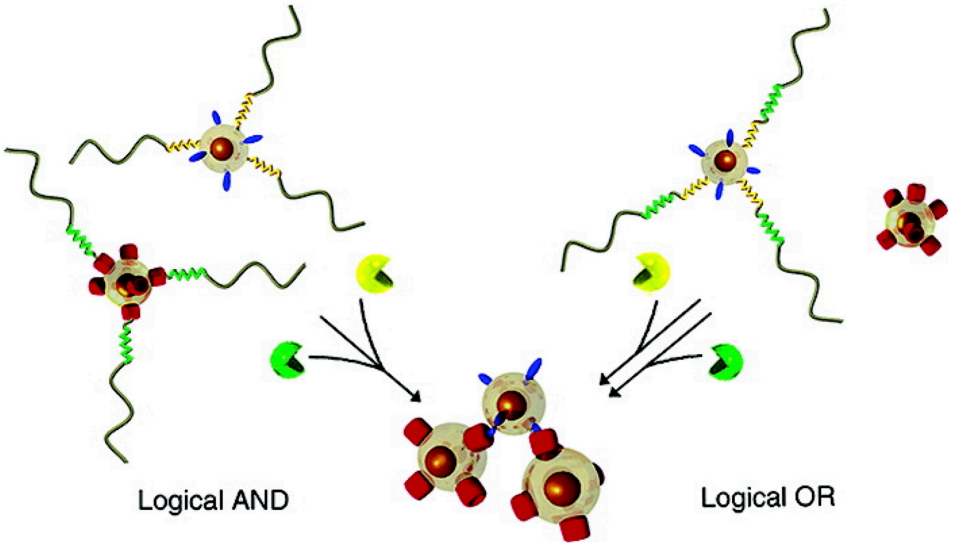



**Figure 1.9:** Nanoparticles functionalized with MMP-cleavable peptides can be used to detect MMP activity under defined logic gates. When the peptides are digested by the required combination of MMPs (yellow and green pacmans), binding sites are exposed, allowing the particles to aggregate and trigger an optical change. Reprinted with permission from [129] - Copyright 2007 American Chemical Society.

### 1.3.3 System disassembly as means of integrating inputs

Biocomputational nanomaterials can also react to input stimuli by either multi-unit nanoparticle disassembly or multi-nanoparticle disaggregation. These platforms often depend on the chemical changes of polymers to induce "system disassembly" [130]. Such disassembly-based platforms have been used primarily for selective targeting of cancer cells for gene therapy and drug delivery. Low pH is one of the most common single inputs for biocomputing nanoplatforms. One report developed a polymer matrix [composed of polyethylene glycol (PEG) and polyethylenimine (PEI)] encapsulating the adeno-associated virus such that upon detection of low pH, the viral vectors are released through triggered chemical expansion [131]. This pH-responsive virus delivery platform may be useful for delivering genes to sites of tumor acidosis. Another interesting platform uses assembly of temperature-responsive block copolymer micelles for drug release [132]. The copolymers have differing low critical solution temperatures (LCST) on each end; thus, depending on the temperature used during particle synthesis, two different types of micelles with differing amounts of loaded cargo can be generated. Each micelle type disassembles under the corresponding temperature stimulus, releasing its drug cargo.

In addition to single input systems, several nanoplatforms have been designed to detect multiple stimuli. One popular approach relies on polymeric micelles that disassemble under the dual stimulus of reducing environment and low pH [133–135]. The micelles encapsulate the drug of interest using polymers that contain both acid-sensitive segments and disulfide bonds. Upon entering an acidic and reducing environment (AND gate inputs), the drug payload is released due to polymer degradation and micelle disassembly (Figure 1.10). Presence of just one of the inputs is not sufficient to rapidly and completely release the cargo.

Other types of AND gate nanoplatforms require different input combinations for system disassembly. Low pH and high temperature is another common pair of AND gate input stimuli for nanoparticles aimed at targeting drugs to areas of inflammation or cancer. Both polymer micelles and dendrimers have been

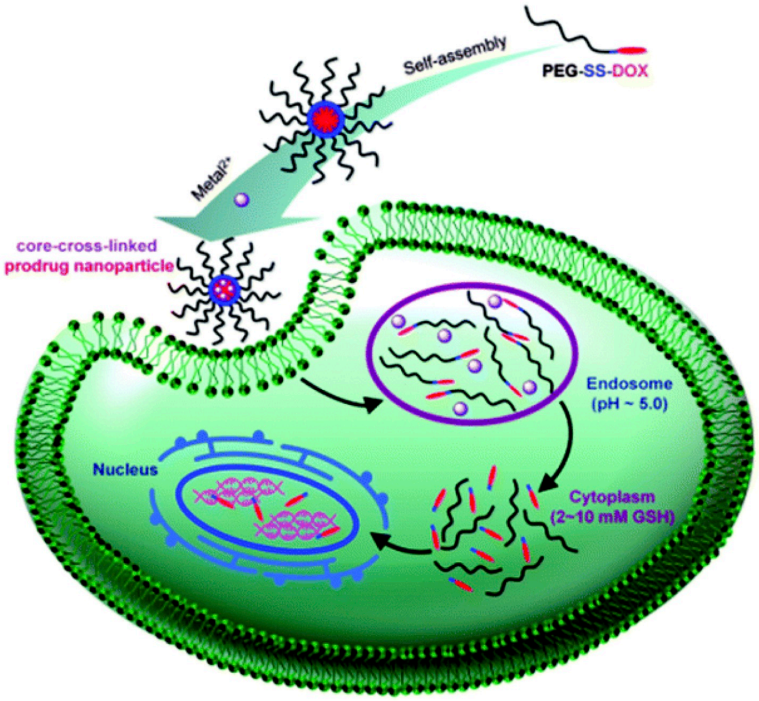



**Figure 1.10:** A self-assembling nanoparticle undergoes AND-gate mediated disassembly and drug release during intracellular trafficking. First, acidic pH in the endosome induces nanoparticle disassembly, then the reducing environment of the cytoplasm breaks the disulfide bond linking the cancer drug doxorubicin (DOX) to polyethylene glycol (PEG). Reprinted with permission from [135].

developed to function as AND gates requiring both low pH and high temperature, two features observed in the local environment of the target tissues [136, 137]. Another similar AND gate system requires a reducing environment as well as UV irradiation for drug delivery [138]. The photosensitive component, a functional nitrobenzyl derivative capable of conformational changes under specific wavelengths of light, is conjugated onto PEG chains that are cross-linked with disulfide bonds. Detection of both UV light by the nitrobenzyl derivative and reduction of the disulfide bonds in the PEG network are required for complete disassembly and release of the encapsulated drug. A more complex example of a multi-stimuli nanoparticle requires pH, temperature, or light to trigger release (i.e. three-input OR gate) [139]. The PEG-based polymeric structure in this platform was designed such that any of these three stimuli is sufficient to trigger disassembly. These reports highlight the potential for functionalized nanoparticles to process many varied stimuli in order to increase targeting specificity. Disassembly-based nanoparticles for drug delivery offer both payload protection throughout transport as well as targeted release.

In addition to therapeutic applications, system disassembly as a method for biocomputing stimuli can be exploited for imaging or diagnostic nanoparticle development. Imaging and diagnostic tools rely on both accurate recognition of the desired target and a detectable visual response. This can be done through changes in fluorescence or absorbance of the nanomaterial-containing solution. For example, the coordination of nucleotide-lanthanide compounds and porphyrin, resulting in a strong fluorescence signal, can be destabilized in the presence of either EDTA or extreme pH values (both low and high) [140, 141]. Particle disassembly takes place through ionic porphyrin displacement, altering the fluorescence properties of the complex and releasing its components. The intensity of this fluorescence signal can be regulated by the concentration and type of chemical inputs. This type of nanoplatform may be further engineered to become useful as a

simultaneous detection and triggered delivery device.

More complex biocomputing nanoplatforms have been developed for the detection of small molecules. One group is interested in developing *in vitro* diagnostics to detect the level of metabolic substrates by generating pH-sensitive polymer-functionalized gold nanoparticles that interface with two cellular enzymes [142]. The enzymes, invertase and glucose oxidase, convert sucrose into acidic byproducts, thereby decreasing the solution pH. The low pH then causes the pH-responsive nanoparticles to undergo multi-nanoparticle disaggregation, resulting in observable absorbance changes. Therefore, this two-tiered information processing platform pairs an AND gate enzyme system with single input (low pH) nanoparticles to yield a visible absorbance change in the solution. Similarly, enzymes covalently attached to the exterior of silicon microparticles have been used to create a local pH change and induce pH-responsive nanoparticle dissolution upon the detection of metabolic substrates in solution [143]. This approach was used to create AND gates responsive to sucrose and oxygen, and OR gates responsive to glucose and ethyl butyrate. Such biocomputing nanoparticles may be used, for example, to image metabolic activity in tissues by quantifying glucose and sucrose availability. Another report developed what could be described as an AND-NOT gated gold nanoparticle platform to detect the simultaneous presence of melamine and the absence of mercury ions [144]. Due to an innate property of gold nanoparticles, the presence of melamine induces the nanoparticles to form multi-nanoparticle aggregates resulting in a blue-colored solution. In the presence of mercury ions, however, the melamine binding sites are blocked by the mercury ions, resulting in a red-colored solution.

In summary, a number of biocomputing nanoplatforms utilize system disassembly, either the disaggregation of multi-nanoparticle structures or the disassembly of nanomaterial components, as the means for integrating stimuli and producing a functional output. The most popular output is the release of therapeutics at target locations, followed by the generation of a detectable fluorescence or absorbance signal for biomedical imaging or diagnostics applications.

### 1.3.4 System transformation as means of integrating inputs

Some biocomputing nanoplatforms undergo a more nuanced change upon stimuli detection and are referred to in this review as undergoing a "system transformation" in response to inputs. A promising approach uses DNA segments that can undergo conformational changes. One system is based on DNA-based "nanorobots" with DNA "locks" that hold the box-shaped nanoplatform closed [145]. Each nanoconstruct can have two different DNA locks, which are aptamers binding to specific target cell-associated antigens, allowing for an AND gate operation. When activated, the nanoconstruct can reveal a variety of payloads, including therapeutic antibody fragments. DNA fragments attached to the surface of gold nanoparticles can utilize system transformation, rather than system assembly/disassembly, to integrate input stimuli. One nanoplatform uses DNA strands to attach fluorophore-quencher pairs to the surface of gold nanoparticles [146]. Upon detection of target DNA sequences in cells, the platform uses toehold displacement to process inputs by altering a fluorophore's proximity to the quenching gold nanoparticle to induce or quench fluorescence. Conformational changes of DNA structures are a powerful tool for controlling the mechanical availability and proximity of nanoparticle components to create sophisticated downstream behaviors.

Another example of using system transformation to integrate biomolecular information is the incorporation of MMP-cleavable masking motifs on the surface of nanoplatforms. For example, masking motifs have been genetically incorporated into the cell surface receptor binding pocket of adeno-associated virus (AAV) nanoparticles such that when the motifs are in place, the vector cannot transduce cells [147]. This system can act as an AND gate by combining two different masking motifs, each susceptible to different MMPs, on the same virus capsid (Figure 1.11). In order for the AAV nanoparticles to successfully transduce

cells, both types of masking motifs need to be cleaved from the capsid. Moreover, the protease-activatable virus nanoparticle was modeled mathematically as a multi-input node, enabling its functional output to be quantitatively predictable. Protease detection systems are particularly promising candidates for targeted disease therapy since MMPs are often upregulated in a variety of disease sites [148].

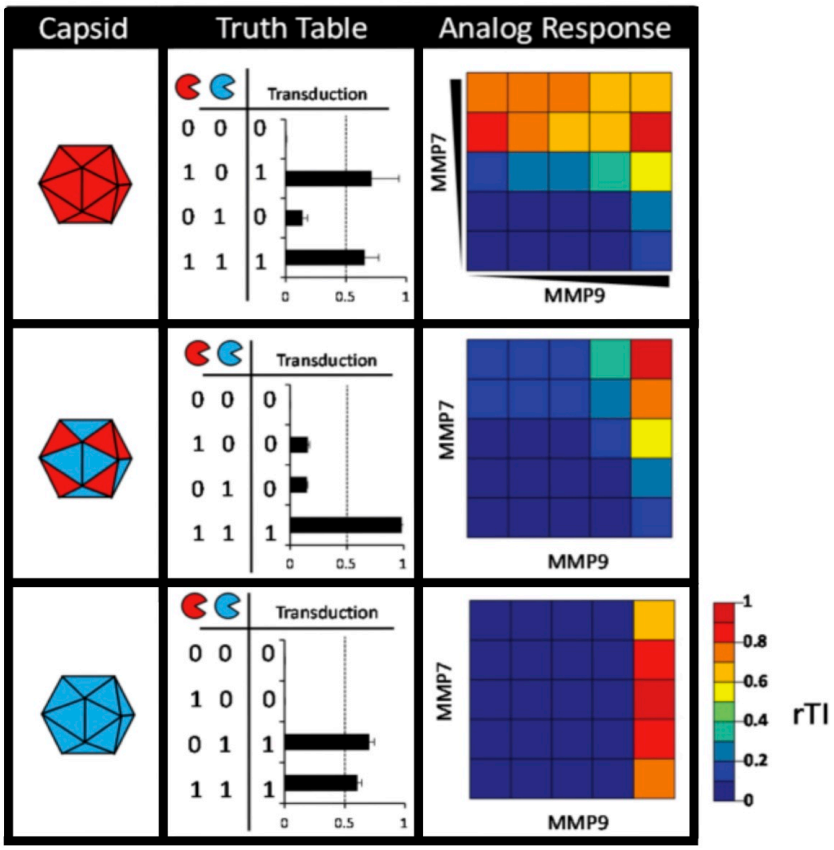



**Figure 1.11:** Virus nanoparticles designed to be activated by two different MMPs for controlled gene delivery. Tunable virus nanoparticles are generated from the combination of two capsid proteins with different masking motifs (red or blue in schematic), each sensitive to a different MMP (red or blue pacmans). The mixed "mosaic capsid" (middle row) demonstrates an AND-gate response to stimulation with the two different MMPs. The "analog response" column displays heatmaps of the relative transduction index (rTI - the percent of fluorescence positive cells multiplied by the mean fluorescence intensity, normalized to control) of the given capsids in response to various concentration combinations of MMP inputs. Adapted with permission from [147].

Nanoparticles have also used various biologically derived materials which exhibit altered physical behaviors, such as motility or swelling, in response to external stimuli. One such system uses two molecular motors to mechanically induce the release of a payload from a hollow nanoparticle for targeted drug delivery [149]. The platform combines light-activated "nanoimpellers" and pH-activated "nanovalves" into an AND gate configuration that requires detection of both light and pH stimuli to allow the payload to escape the nanoparticle. Another system transformation-based approach uses phage capsid proteins, which undergo a structural change upon binding to their unique environmental targets [150]. Two different phages can be combined into one phage-bundle nanostructure to create an AND gate. Thus, detection of two different nitroaromatic molecules results in swelling of the nanostructure, leading to a change in reflected color based

on the arrangement of fibers. Thus, these phage nanostructures can be used to detect the presence of explosives by identifying multiple signature chemicals. In summary, biocomputing nanoplatforms that utilize system transformation harness a more subtle change of the nanomaterial as the means for sensing inputs and generating outputs. Many of these platforms are built from naturally occurring materials, such as DNA, viruses, and proteins.

### 1.3.5 Conclusions and perspectives

The future of successful therapeutic and diagnostic nanoplatforms hinges on new and improved mechanisms for connecting inputs (both intrinsic and extrinsic) to predictable and reliable output responses. The wide range of platforms discussed here represents a promising foundation to achieving this vision. System assembly-based biocomputing nanoplatforms show promise in many diagnostics applications, as generation of multi-nanoparticle aggregates can lead to changes in absorbance or fluorescence of the nanoparticle solution. Nanoparticle aggregation or multi-subunit assembly allows for signal integration through the coming together of many signal-detecting components. Biocomputing platforms based on system disassembly have shown more promise in the field of drug delivery. Multi-nanoparticle disaggregation or multi-subunit disassembly allows for the release of therapeutic cargo only at the site of input detection. Finally, system transformation uses more nuanced changes to the nanomaterials to yield desired functional outcomes, and shows promise in imaging and diagnostics, as well as for drug delivery.

In order for more sophisticated multi-tiered algorithmic trees to be built from individual nanoparticle logic gates, the individual nanoparticles should be better characterized quantitatively. Based on experimental data, the "transfer function" of each nanoparticle should be determined under various parameters. These quantitative functions can help predict how a system would function given a certain set of conditions. Thus, approaching nanoparticle design with quantitative biocomputation in mind should lead to the generation of more predictable nanoplatforms.

Additionally, if therapeutic biocomputing nanoplatforms are to be used in complicated biological contexts, such as inside cells or in intact animals, the concentrations of various input stimuli need to be quantified in relevant biological environments, such as diseased cells and tissues. This information is being curated by various databases, but patient-to-patient variability may require personalized measurements of the target biomarker(s) either via biopsies or through the use of companion diagnostics. The nanoplatforms must then be optimized for the gradients of stimuli they will encounter in the target biological environments. In conclusion, the field of biocomputing nanomaterials is expanding to encompass many more complex systems than ever before. By conceptualizing the design process as one analogous to building electrical circuits, engineered nanomaterials may be programmed to carry out sophisticated, multi-step computations that can result in better targeted and more efficient drugs, or the detection of complex mixtures of desired analytes.

## 1.4 Strategies for engineering stimulus-response into protein-based components

Protein engineering strategies are crucial for the modification of genetically encoded nanoparticles to create new functionality. As improved technology to study protein interactions has produced a better understanding of their binding interfaces and dynamic behaviors, researchers have capitalized on this knowledge to alter proteins for medical and manufacturing purposes. Protein engineering has primarily focused on changing enzyme activity to modify biochemical pathways [151]. Molecular dynamics models have been applied to

alter existing proteins and create novel enzymes. In addition, libraries of mutated proteins have been built and screened to identify useful variants.

### 1.4.1 Rational design of biological behaviors

The rational design approach to modifying proteins analyzes structures and identifies likely key residues by their proximity to interaction domains, homology in proteins with similar structure, and chemical and physical properties [151]. Then, these residues are modified to increase or decrease binding interaction strength, alter hydrophobic effects, sterically block interactions, or modify protein flexibility by altering ring structure [151]. Rational design has also been used to design enzymes with completely new functions by starting with a generic base tertiary structure or an existing enzyme and using computational tools to create a new interaction domain [152]. Computational models can identify binding residues on a ligand and create a pocket to interact with these residues in the correct shape, based on existing structures [153]. This approach is then applied to create binding pockets that act as enzymes by facilitating the stability of intermediate states (Figure 1.12) [154, 155]. In addition to these strategies for creating novel enzymes, bioinformatic approaches compile information from existing catalytic domains to identify regions necessary for function and regions likely susceptible to mutation for retargeting [156].

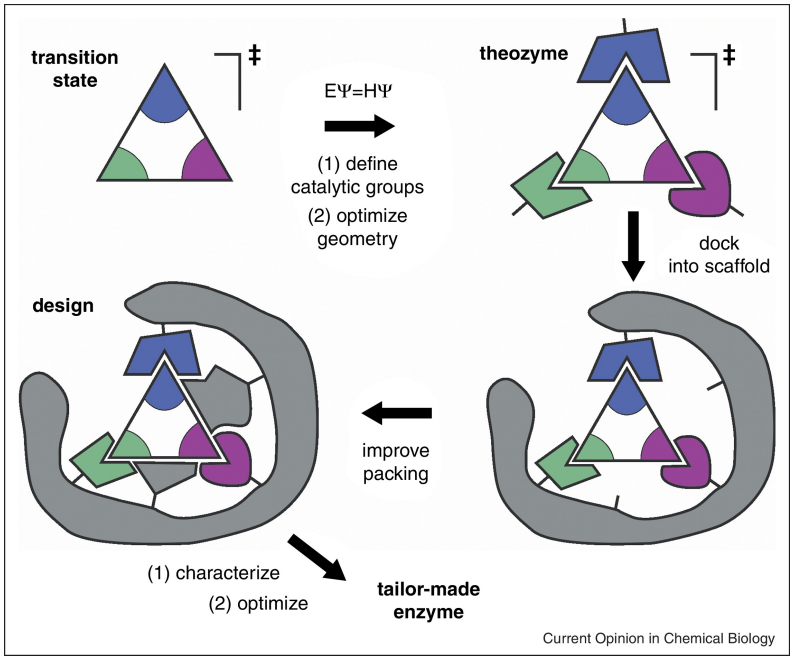



**Figure 1.12:** A computational approach for designing enzymes that stabilize transition states. Reprinted with permission from [154].

### 1.4.2 Library-based evolution of dynamics

Another approach to creating proteins with modified dynamics is to create a mutation library from an existing protein and then select for desired behaviors. While libraries can be generated through random mutation, typically the mutation is guided by principles similar to those used in rational design approaches. Mutational libraries can be directed at structurally significant regions, or regions identified as susceptible to mutation through evolutionary analysis [157, 158]. These libraries can be created through point and short sequence

mutations or through recombination. Screens for these libraries are tricky to develop, as they must both isolate the target dynamic property and preserve sequence information. Thus, these studies are either cell based, relying on cell sorting or high thoroughput approaches, or incorporate unique peptide labels on each mutant [159]. Evolution-based approaches iterate through library generation and screening, incorporating computational evaluation of successful mutants to optimize future generations (Figure 1.13).

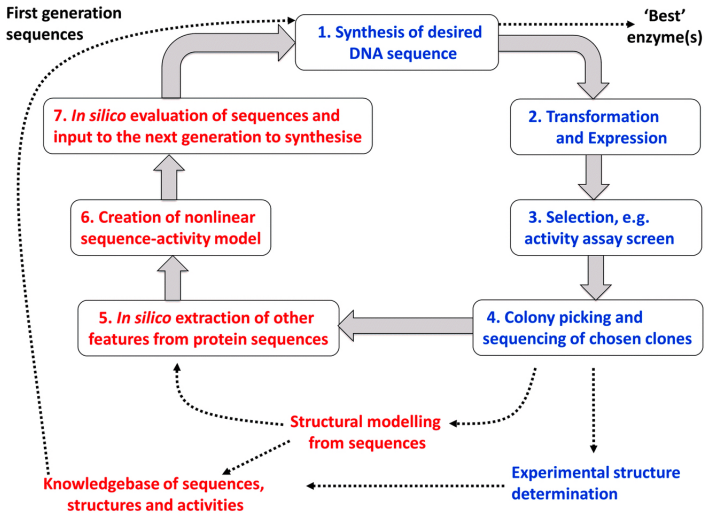



**Figure 1.13:** A computational approach for iterating libraries to generate optimized enzymes. Reprinted with permission from [160].

## 1.5 Thesis overview

This thesis describes the development of two technologies for peptide and protein display on the AAV capsid, as well as the application of computational design tools for accelerating the process of engineered viral capsid development. Chapter 1 introduces AAV as an engineerable nanoplatform with therapeutic relevance, summarizes previous research on biocomputing nanoplatforms, and describes techniques for engineering stimulus-response behaviors into protein-based systems. Chapter 2 describes the development of a panel of AAV-based viral nanoparticles capable of varying degrees of constant or activatable peptide display in response to heat stimulus, derived from an innate-to-AAV peptide-display function. Chapter 3 demonstrates a different approach for displaying larger proteins or targeting motifs on the AAV capsid by attaching them post-assembly, and explores potential applications for the delivery of gene editing machinery. Chapter 4 examines two computational approaches, one derived from coarse-grained molecular dynamics modeling and one drawn from the study of residue co-evolution within protein families, for their ability to predict mutant AAV capsid formation and function. Overall, this thesis expands the potential outputs of engineered AAV platforms and presents strategies for applying computational models to speed the viral nanoplatform development pipeline.

# Chapter 2

# Reprogramming the Activatable Peptide Display Function of Adeno-Associated Virus Nanoparticles

†

We harnessed an intrinsic activatable peptide display behavior shared by several parvoviruses, including the adeno-associated virus (AAV), in order to design protein-based nanodevices that can carry out an exogenous functional output in response to stimulus detection. Specifically, we generated truncated viral capsid subunits that, when combined with native capsid components into mosaic capsids, can perform robust activatable peptide display. By modulating the ratio of subunits in the mosaic capsid, properties of the activatable peptide display function can be optimized. Interestingly, the truncated subunits can form homomeric capsids not observed in nature, but at the price of losing the ability to carry out activatable peptide display. Collectively, our results demonstrate the importance of capsid mosaicism when activatable peptide display is desired, and help explain why the wild-type AAV capsid exists as a mosaic of different subunits. This proof-of-concept study illustrates a strategy for reprogramming a particular conformational output behavior of AAV in pursuit of the long term vision of creating stimulus-responsive nanodevices.

## 2.1 Introduction

Viruses are self-assembling nanodevices whose capsids exhibit energetic metastability, allowing for conformational shifts in response to environmental stimuli. Due to their innate ability to protect and deliver biomolecular and genomic cargo, viral nanoparticles (VNPs) are being developed as diagnostics and therapeutics [14,161]. VNPs exhibit promise for predictable design and engineering due to their genetic encoding, monodispersity, and controlled self-assembly. These traits have been leveraged to develop 'bionic' viruses, integrating natural and synthetic components to improve the specificity and efficiency of viral infection, as well as the range of therapeutic and diagnostic functional outputs [15].

In order to gain greater control over viral function and to expand viral capabilities for sensing and responding to desired environmental cues, the field of synthetic virology seeks to 1) identify, characterize, and

†This chapter has been adapted with permission from Thadani, N. N., *et al.* (2018). Reprogramming the Activatable Peptide Display Function of Adeno-Associated Virus Nanoparticles. *ACS Nano.*

refactor viral elements and 2) reprogram the intrinsic 'inputs' and 'outputs' of viruses. The latter strategy, broadly termed biocomputation, allows nanoplatforms to detect, integrate, and process environmental information and to produce predictable outputs in response [162]. Ideally, the design of these computing nanoparticles is modular, allowing for different input-sensing domains to be combined with various output-producing components, resulting in complex signal integration. This strategy applied to VNPs may further augment their innate aptitude for cellular infection, enabling improved specificity and efficiency of tissue targeting, disease diagnosis, and cargo delivery.

Viral capsid engineering has generally focused on modifying the 'inputs' governing infection. For example, viral components have been mutated to detect desired host cell-specific receptors as inputs, thus altering cellular binding [163,164]. VNPs have also been designed to detect other types of inputs, such as extracellular enzymes to trigger viral cell entry [147] and exogenously applied light to alter viral intracellular trafficking [165]. Magnetic and pH-responsive VNP systems have been developed by conjugating viruses to materials such as magnetic metal nanoparticles [166] or pH-responsive peptide matrices, [131, 167] translating the stimulus-responsive properties of these materials to viral delivery. Upon detection of input stimuli, the most common output produced has been delivery of the cargo carried by the VNP. Other than modifying what cargo is carried, strategies for reprogramming other types of VNP outputs have been largely unexplored.

As proof-of-concept, we sought to reprogram a dynamic conformation change 'output' present in a virus capsid. Specifically, upon entry into a host cell's endosomal pathway, the capsid of several parvoviruses, including that of adeno-associated virus (AAV), undergoes a structural shift leading to the surface-display of previously hidden peptide motifs [69,168–171]. AAV is a small 25 nm diameter virus consisting of a single-stranded DNA genome and a 60-mer mosaic capsid composed of three protein subunits (VP1, VP2, and VP3 in a 1:1:10 stoichiometric ratio). These VPs are encoded on a single open reading frame (ORF) of the AAV *cap* gene with unique start sites, resulting in capsid subunit proteins that share a common C-terminal domain but have progressively fewer N-terminal residues (Figure 2.1). The longer N-terminal regions of VP1 and VP2 are presumed to be packaged in the interior of the AAV capsid, then externalized onto the capsid surface during endosomal trafficking due to the low pH and presence of other endosomal factors [9]. The N-terminus of VP1 contains a phospholipase A2 (PLA2) domain and putative nuclear localization signals (NLSs) that facilitate endosomal escape and nuclear translocation, respectively [70, 71]. In other words, the AAV capsid exhibits a stimulus-activatable peptide display functionality. This peptide externalization is thought to involve the capsid five-fold pores, although the two-fold axis of symmetry may also undergo a conformational change during the process [43, 72, 73]. The activatable peptide display behavior may be induced experimentally by heating the capsid to ~62 °C. We previously demonstrated successful reprogramming of this structural output by replacing the PLA2 domain (ΔPLA mutant) with a hexahistidine ($His_6$) tag, which enabled the virus to bind nickel ions upon heat activation [172].

Here, we built upon our prior work to investigate the design rules foundational to the construction of virus-based nanodevices that can use this type of capsid conformational switching in order to output defined functions. Interestingly, AAV's VP1 and VP2 subunits are able to carry out activatable peptide display; however, VP1 and VP2 subunits cannot form homomeric capsids that are composed of a single subunit type (*e.g.* entirely VP1 subunits or entirely VP2 subunits). On the other hand, the shortest VP3 subunit can form homomeric capsids (entirely VP3 subunits) but does not exhibit activatable peptide display behavior [173]. Thus, we rationalized that an engineered capsid subunit with a length between that of VP2 and VP3 may be able to form homomeric capsids and exhibit dynamic peptide display. To test this hypothesis, we generated and characterized a panel of VP2 truncation mutants, and our results demonstrate the importance of capsid mosaicism—the mixture of different subunit types into one capsid—for the proper functioning of AAV's

activatable peptide display.

## 2.2 Results

### 2.2.1 Characterization of homomeric VP2 truncation VNPs

Using site-directed mutagenesis, we generated six VP2 truncation mutants by first silencing the VP1, VP2, and VP3 start codons and then inserting a new start codon in progressively more downstream locations from the original VP2 start, resulting in the deletion of increasingly larger numbers of amino acids from the VP2 N-terminus (Figure 2.1). A hexahistidine tag ($His_6$) immediately follows the inserted start codon, yielding a panel of VP2Δ-$His_6$ mutants.

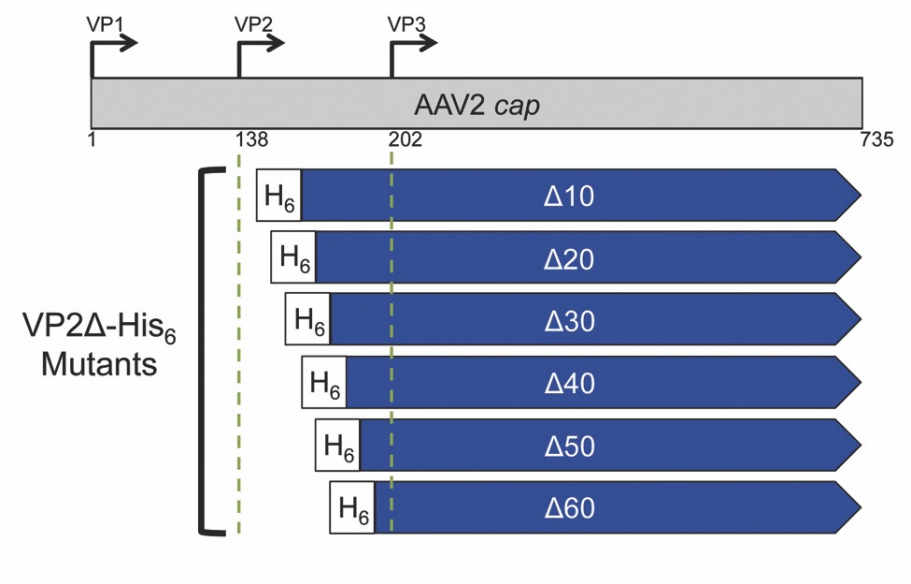



**Figure 2.1:** A schematic of wild-type AAV2 *cap* gene with an open reading frame containing overlapping genes for VP1, VP2, and VP3, and the VP2Δ-$His_6$ mutant genes generated from the *cap* gene. The amino acid residue numbers of the native start codons are indicated below the wt AAV2 gene. VP2Δ-$His_6$ mutant genes consist of a start codon, a $His_6$ sequence, and serial 10-AA truncations of the VP2 protein sequence. Schematic not drawn to scale.

The VP2Δ-$His_6$ mutants were used to form homomeric capsids, each containing only identical copies of a single VP2Δ-$His_6$ protein. Surprisingly, all mutants were able to assemble homomeric capsids and package genomes, as evidenced by quantitative PCR (qPCR) to measure viral titers and western blotting to detect capsid subunits (Figure 2.2a, c). These titers appear lower than the ΔPLA mutant (used here as a control, as it forms with titers similar to wt), but the differences are not statistically significant. All VNPs have significantly lower titers than homomeric VP3 VNPs. In comparison, the wt VP2 subunit with the full-length VP2 N-terminus cannot form homomeric capsids (data not shown), as previously reported by others [173]. Viral titers have an inverse relationship with the length of the VP2Δ-$His_6$ mutants, with the exception of Δ60, whose low titer excluded it from further study.

Although the VNPs are able to assemble into capsids that contain genomes (Figure 2.2a), the structural integrity or morphology of the capsids may be compromised by the incorporation of 60 VP2 N-termini, as opposed to the ~10 N-termini in wt capsids (5 VP1 and 5 VP2 N-termini per wt capsid). We tested

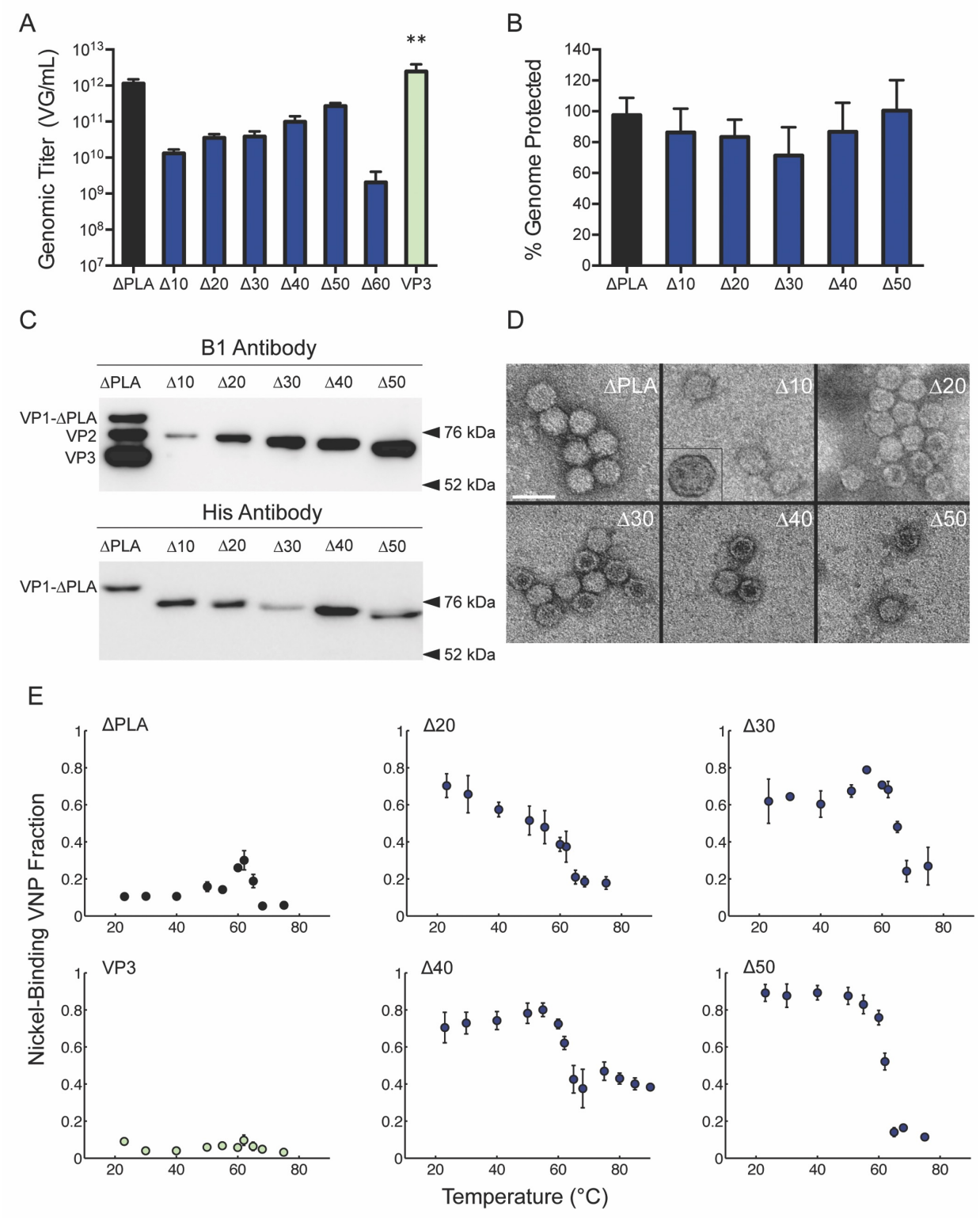



**Figure 2.2:** Characterization of homomeric VP2Δ-$His_6$ virus nanoparticles (VNPs). Caption continued on following page.

**Figure 2.2** ***(previous page)*****:** A: Genomic titers of 10-plate virus preps determined using qPCR and expressed as viral genomes per mL (Δ20 N=6, Δ30 N=3, Δ50 N=4, all others N=2). The ΔPLA-$His_6$ VNP (ΔPLA) composed of wtAAV2 proteins with the PLA domain on VP1 replaced with a $His_6$ domain (previously characterized [172]) and a VNP (VP3) composed solely of the shortest capsid protein VP3 are included as controls. ** $P<0.01$ when compared to all VNPs shown. B: Benzonase genomic protection assay (Δ20 N=6, Δ10 and Δ40 N=3, all others N=4). VNP samples were treated with benzonase or sham buffer and remaining genomes quantified with qPCR. % genome protected is the fraction of genomes in the benzonase-treated sample as compared to sham. The Δ60 mutant was excluded from the genomic protection assay and other further assays due to low titer. C: Western blots of homomeric capsids. Assembled capsids extracted from cell lysates and separated on iodixanol gradients were denatured before gel separation and western blotting using a B1 antibody (top) and an anti-His antibody (bottom). The B1 antibody recognizes a C-terminal epitope present in all VPs. D: TEM images of homomeric capsids. Empty capsids appear as hexagonal shapes with dark centers, while full capsids appear as hexagonal shapes of uniform color. 150,000X, scale bar is 50 nm. E: Temperature responsive nickel binding assays of homomeric capsids (ΔPLA and Δ40 N=4, all others N=3). Viruses were incubated at various temperatures, applied to nickel affinity columns, and the amount of virus bound to the column was quantified using qPCR. The Δ10 and Δ60 were excluded due to insufficient titers. All error bars are SEM, and statistical comparisons were conducted using ANOVA with post-hoc testing.

the abilities of the VNPs to protect their genomes from nuclease digestion by incubating the particles with benzonase and using qPCR to quantify the number of viral genomes protected by a structurally intact capsid (Figure 2.2b). Genome protection is comparable to the ΔPLA control for all VNPs tested.

We qualitatively assessed the morphology of the virus variants using transmission electron microscopy (TEM). The VP2Δ-$His_6$ homomeric capsids display hexagonal morphology similar to that of the ΔPLA-His virus and samples consist of both full and empty capsids (Figure 2.2d). The population of Δ10 mutants also contains a subset of capsid-like structures that are roughly twice the size of a normal AAV capsid (Figure 2.2d, inset); these abnormal structures were not observed for any of the other mutants. These results thus far demonstrate most of the VP2Δ-$His_6$ homomeric capsids are structurally intact and appear morphologically normal.

To test the activatable peptide display of the VP2Δ-$His_6$ homomeric VNPs, we used nickel affinity chromatography. The VNPs were incubated at a variety of temperatures between 20 °C and 75 °C and then applied to nickel affinity columns. After unbound and denatured VNP genomes were removed through washing, the elution fraction was collected and viral content quantified as a percentage of total viral genomes collected from the column. This analysis was conducted for the Δ20-50 mutants, as the Δ10 and Δ60 homomeric capsids did not form with sufficient titers. As previously demonstrated by Musick *et al.*, [172] the ΔPLA-His VNP exhibits maximum binding at 62 °C—approximately 40% of the capsids bind the column (Figure 2.2e, top left). At higher temperatures the capsid denatures, so no binding of intact capsids is observed.

Interestingly, the VP2Δ-$His_6$ homomeric VNPs exhibit binding profiles strikingly different than the ΔPLA-$His_6$. All VP2 truncation VNPs demonstrate high nickel binding at room temperature (RT) (Figure 2.2e), suggesting that the $His_6$ tags are already surface-exposed on the homomeric capsids. The Δ20 appears to lose column binding ability as incubation temperature increases, at temperatures well below the previously reported ~72 °C melting temperature of wt AAV2 [97]. The Δ30 and Δ40 exhibit a slight gain in column binding with increasing temperature until a peak is reached at ~55 °C, followed by a drop-off. This observation may suggest that a minor subset of $His_6$ tags are activatable (*i.e.* revealed upon temperature activation) for Δ30 and Δ40 homomeric capsids, while the rest are always surface-exposed. Notably, the

Δ40 homomeric capsid appears to bind the nickel column well past the wt denaturation point, as determined by extending the assay to temperatures up to 90 °C. The Δ50 maintains high nickel binding with no detectable change until 60 °C, after which point the binding drops off likely due to capsid denaturation. Generally, the VP2Δ-$His_6$ homomeric capsids exhibit negligible activatable peptide display (as determined by the difference in binding at 60 °C compared to binding at RT), but exhibit a greater maximum avidity for the nickel columns than ΔPLA-$His_6$. These results suggest that the N-terminal domains of the VP2Δ-$His_6$ variants do not pack effectively inside homomeric VNP capsids, and the viruses are in an "ON" state with regards to nickel binding capability irrespective of temperature activation. The binding capabilities mainly decrease due to capsid denaturation.

### 2.2.2 Characterization of mosaic VP2 truncation VNPs

The wt AAV capsid is not homomeric, but rather is a mosaic of three capsid subunits—VP1, VP2, and VP3. To determine the impact of mosaicism on the function of the engineered N-terminal domain, we generated mosaic capsids composed of VP2Δ-$His_6$ mutants and VP3, the shortest wt AAV capsid subunit that does not exhibit activatable behavior. As a first pass, we generated mosaics that theoretically contain one part VP2Δ-$His_6$ and three parts VP3 subunits. In wt capsids VP1, VP2, and VP3 exist in a 1:1:10 ratio, so VP3 makes up ~5/6 of the capsid components. Thus, our first panel of mosaic capsids have slightly less VP3 than in wt AAV. We then characterized the formation, incorporation of components, and structural integrity of these VP2Δ-$His_6$ 1:3 mosaic VNPs.

In contrast to homomeric capsids, all 1:3 mosaic capsids assemble with titers comparable to wt, indicating a recovery in capsid assembly (Figure 2.3a). All mosaic capsids also protect their genomes from nuclease digestion similarly to wt, as indicated by genomic protection assay (Figure 2.3b).

We confirmed that the 1:3 mosaic capsids incorporate both VP2Δ-$His_6$ and VP3 subunits by B1 western blot, and the anti-$His_6$ western blot verified inclusion of $His_6$ in the truncated VP2 subunits (Figure 2.3c, d). The B1 western indicates successful incorporation of all VP2Δ-$His_6$ mutant subunits in their respective mosaic capsids except Δ50 and Δ60, which are too close in molecular weight to VP3 to be discerned. However, all VP2Δ-$His_6$ mutant subunits can be identified in their respective mosaic VNPs on the anti-$His_6$ blot.

Nickel affinity chromatography of the VP2Δ-$His_6$ 1:3 mosaic VNPs reveal restored activatable peptide display for almost all variants (Figure 2.3e). The VNPs exhibiting activation reach peak column binding at 60 °C. Peak binding ranges from 63% ($\Delta 20_1$-$VP3_3$) to 75% ($\Delta 60_1$-$VP3_3$), an increase over the 40% previously shown for ΔPLA-$His_6$ VNP (Figure 2.2e). Binding at RT varies among the mosaic VNPs, from 13–20% for longer VP2Δ-$His_6$ mutants to 30% for shorter mutants, suggesting longer N-termini are better at packaging and concealing the $His_6$ motif. Interestingly, the $\Delta 40_1$-$VP3_3$ mosaic VNP does not exhibit activatable binding—rather, its phenotype resembles the homomeric VNPs, with high column binding at RT and a drop-off upon denaturation. This may be indicative of deficient N-terminal packaging of the Δ40 VP2Δ-$His_6$ subunits in the 1:3 mosaic capsid. All of the mosaics largely appear to denature by 70 °C. However, the $\Delta 10_1$-$VP3_3$ mosaic shows significant column binding at 70 °C as previously seen in the Δ40 homomeric VNP, while the $\Delta 40_1$-$VP3_3$ does not exhibit column binding at 70 °C. Taken together, mosaicism of the virus capsid appears to be an important design strategy for achieving temperature-activatable N-terminus externalization.

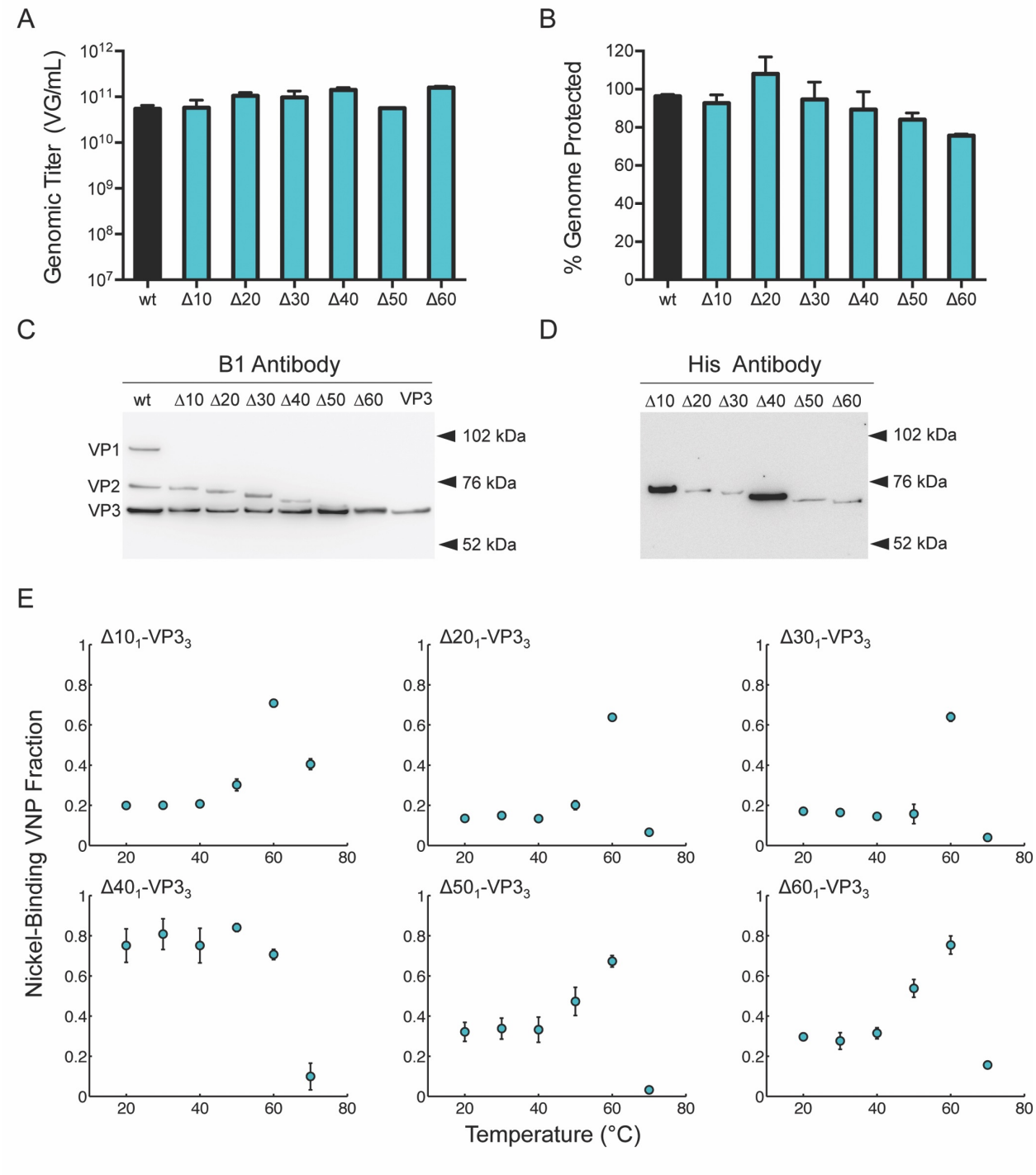



**Figure 2.3:** Characterization of VP2Δ-$His_6$ 1:3 mosaic VNPs. Caption continued on following page.

**Figure 2.3** ***(previous page)*****:** A: Genomic titers of one-plate virus preps determined using qPCR and expressed as viral genomes per mL (Δ20 N=3, all others N=2). wtAAV2 included as a control. B: Benzonase genomic protection assay (Δ30 N=6, all others N=3). VNP samples were treated with benzonase or sham and remaining genomes quantified with qPCR. % genome protected is the fraction of genomes in the benzonase-treated sample as compared to sham. C: Western blot of the 1:3 mosaic capsids. Assembled capsids extracted from cell lysate and separated on iodixanol gradients were denatured before gel separation and western blotting using a B1 antibody. D: Western blot of the 1:3 mosaic capsids as described in C, using an anti-His antibody (image intensity adjusted uniformly). E: Temperature responsive nickel binding assays of the 1:3 mosaic capsids (N=3). Viruses were incubated at various temperatures, applied to nickel affinity columns, and the amount of virus bound to the column was quantified using qPCR. All error bars are SEM, and statistical comparisons were conducted using ANOVA.

### 2.2.3 Impact of subunit ratio on mosaic VNPs

Results from the homomeric and 1:3 mosaic VNPs suggest that the ratio of VP2Δ-$His_6$ to VP3 may play a role in activatable N-terminus externalization. To explore this relationship further, we developed a panel of Δ30:VP3 mosaics at different ratios, spanning from 3:1 to 1:5 with the latter being most similar to the ratio of VP1 and VP2 subunits to VP3 in wt AAV2. We then characterized this Δ30 mosaic panel to determine the impact of different subunit ratios on VNP formation, subunit incorporation, structural integrity, and temperature-responsive nickel binding.

Δ30 mosaic VNPs at all ratios assemble and protect their genomes from nuclease digestion similar to wt. (Figure 2.4a, b). To determine if altering the transfection plasmid ratios of Δ30:VP3 yields VNPs with the expected ratio of subunits in assembled capsids, we conducted B1 western blots (example, Figure 2.4c) and compared the intensities of capsid subunit bands within each lane using densitometry (Figure 2.4d). Ratios of proteins in the assembled VNPs correspond to transfection ratios ±15%, indicating that transfection ratios may be used to control mosaic ratios in aggregate populations of VNPs, although mosaic incorporation ratios may vary between individual capsids.

Nickel affinity chromatography of the Δ30:VP3 mosaic VNPs indicates that the ratio of Δ30 to VP3 subunits impacts RT binding as well as the degree of activatable behavior (Figure 2.4e). The 3:1 VNP exhibits a phenotype similar to the homomeric VNPs, with high column binding at RT and a drop-off upon denaturation. The 1:1 VNP exhibits a phenotype of partial RT binding (58%), and activation to a peak of 83% binding—the highest observed of any VNP exhibiting activation. The 1:3 VNP (previously described in Figure 2.3) and 1:5 VNP exhibit low RT binding and peak activation around 60 °C. Thus, for mosaic VNPs, a greater proportion of VP3 in the capsid results in less RT binding and greater activation. Collectively, these results suggest that activatable subunits should be the minor component of mosaic VNPs if robust activatable N-terminus externalization is desired.

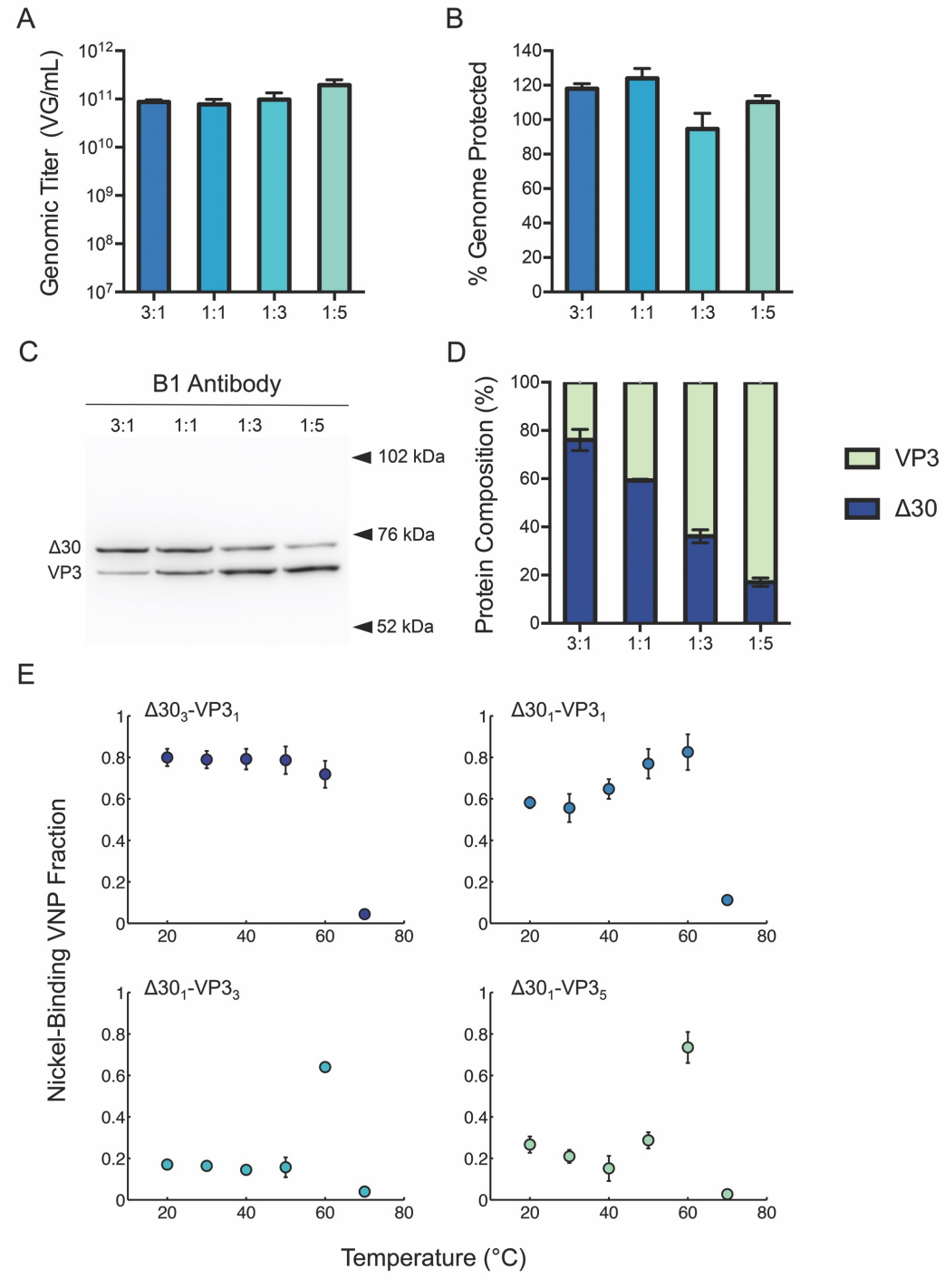



**Figure 2.4:** Characterization of Δ30:VP3 mosaic VNPs. Caption continued on following page.

**Figure 2.4** ***(previous page)*****:** A: Genomic titers of one-plate preps determined using qPCR and expressed as viral genomes per mL (N=2). B: Benzonase genomic protection assay (1:3 mosaic N=6, all others N=3). VNP samples were treated with benzonase or sham buffer and remaining genomes quantified with qPCR. % genome protected is the fraction of genomes in the benzonase-exposed sample as compared to the sham-exposed sample. C: B1 western blot of the Δ30 mosaic capsids. Assembled capsids extracted from cell lysates and separated on iodixanol gradients were denatured before gel separation and western blotting using a B1 antibody. D: Subunit composition of Δ30 mosaic VNPs determined through densitometry of B1 western blots (N=2). E: Temperature responsive nickel binding assays of the Δ30 mosaic capsids (N=3). Viruses were incubated at various temperatures, applied to nickel affinity columns, and the amount of virus bound to the column was quantified using qPCR. All error bars are SEM, and statistical comparisons were conducted using ANOVA.

## 2.3 Discussion

To understand the relationship between VNP composition and activatable peptide display behavior, we analyzed the impact of VNP mosaic components and ratios on capsid assembly and activation. We defined an Activation Index as the difference in binding after incubation at peak activation temperature (60–62 °C) and binding after RT incubation (20–23 °C).

Capsid assembly in general is not significantly correlated with VP2Δ-$His_6$ truncation mutant length (Table 2.1); however, among the homomeric capsids, shorter truncation mutants appear to favor assembly (Figure 2.2a). This trend is broken by the Δ60 capsid. The Δ40, Δ50, and Δ60 sequences introduce mutations into the assembly-activating protein (AAP), a protein encoded in an alternate reading frame on the *cap* gene [6]. AAP is required for assembly of AAV2, although this co-factor is not required for the assembly of all AAV serotypes [31]. While the Δ40 and Δ50 protein sequence mutations introduce a short peptide sequence (CYYYYYY) to the N-terminal end of AAP, the Δ60 mutation introduces a stop codon. This mutation may block the expression of functional AAP, resulting in failed assembly of the homomeric Δ60. When Δ60 is assembled in mosaic capsids with VP3 subunits, AAP is produced from the VP3-expressing plasmid, potentially accounting for restored capsid assembly. In general, capsid assembly is positively correlated with increasing amounts of VP3 (Table 2.1). This could be due to VP2Δ-$His_6$ truncation mutants promoting increased formation of empty capsids or capsid protein oligomers that are separated from full capsids during iodixanol gradient ultracentrifugation [21]. VP3-only capsids form with significantly higher titer than ΔPLA-$His_6$ capsids (with composition similar to wt), consistent with our observation that mosaics with higher amounts of VP3 form more VNPs.

Genomic protection is not significantly correlated with VP2Δ-$His_6$ truncation mutant length or the amount of VP3 incorporated (Table 2.1). VNPs exhibit genomic protection comparable to wt. Nickel column binding assays, however, hint at differences in capsid stability after exposure to high temperatures.

**Table 2.1:** Pearson correlations of VNP subunit composition and VP2Δ-His6 mutant length with capsid assembly, genomic protection and Activation Index (N=13). Activation Index is defined as the difference in binding at the peak activation temperature and at RT. p<0.05:*, p<0.01:**, p<0.001:***

| | **Capsid Assembly** | **Genomic Protection** | **Activation Index** |
|---|---|---|---|
| **VP2Δ-$His_6$ Length** | -0.107 (*n.s.*) | 0.104 (*n.s.*) | 0.083 (*n.s.*) |
| **Amount VP3** | 0.835*** | 0.306 (*n.s.*) | 0.805*** |

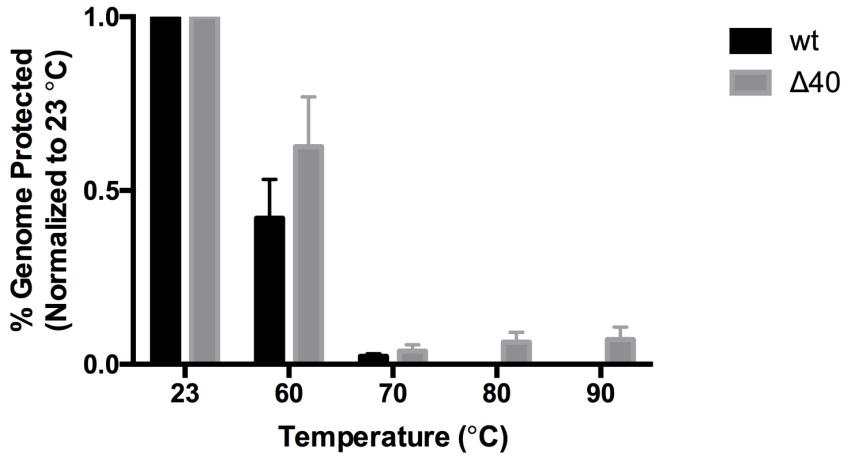



**Figure 2.5:** Benzonase genomic protection assay of Δ40 homomeric VNP post-incubation. Δ40 was incubated at the stated temperatures for 30 minutes, then benzonase genomic protection assay was conducted. wt AAV2 is included as a control. Genomic protection was normalized to the level of protection at 23°C for both VNPs. Δ40 exhibits low degrees of genomic protection after incubation at 80°C and 90°C, while wt exhibits no genomic protection after incubation at these temperatures. Error bars are SEM (N=3).

In particular, the majority of VNPs exhibit a near-complete drop-off in column binding by 70 °C, near the melting temperature of wt at 72 °C [97]. However, the Δ40 homomeric VNP and Δ10 mosaic VNP still bind the column at the highest temperatures tested, up to 90 °C for the Δ40 homomeric VNP. (Figures 2.2e, 2.3e). This sustained binding may be due to the Δ10 and Δ40 truncation mutant proteins possessing increased viral genome affinity. These protein monomers may hold genomes on the columns even at temperatures where the capsid has denatured. Alternatively, these Δ10 and Δ40 mutants may facilitate greater capsid thermal stability. Preliminary data from benzonase assays conducted on the Δ40 homomeric VNP after exposure to high temperatures indicate that a fraction of this VNP exhibits enhanced thermal stability (Figure 2.5). It is unclear why these VNPs exhibit enhanced stability differentially between homomeric and mosaic capsids, and future studies examining the structures of the mutant capsids are needed to further explain these observations.

Activatable peptide display, quantified *via* the Activation Index, is positively correlated with increasing amounts of VP3 but has no significant correlation with the length of VP2Δ-$His_6$ truncation mutant incorporated (Table 2.1). To identify shared characteristics of activatable VNPs, we classified the VNPs using k-means clustering into two clusters on the Activation Index for all VNPs with nickel column data (Figure 2.6a). We then identified the cluster with the higher Activation Index as activatable and the other as non-activatable. This clustering results in an activation threshold of approximately 0.2 and a ratio of sum of squares (ss) of distances between clusters over total sum of squares of distances between all data points of 84%, indicating that 84% of variance in the Activation Index dataset is explained by these clusters. Interestingly, all capsids with a ratio of VP2Δ-$His_6$:VP3 less than or equal to 1:1 (*i.e.* same or less proportion of VP2 truncation subunit in mosaic relative to VP3) are classified as activatable, with the exception of the $\Delta40_1$-$VP3_3$ mosaic capsid. It is possible that the proline-rich region near the Δ40 N-terminus contributes to structural rigidity, making it difficult for the N-terminal $His_6$ tag to package inside the capsid for later activation [174]. Further investigation is needed to elucidate the properties of Δ40 preventing activatable peptide display, even in a majority-VP3 mosaic capsid.

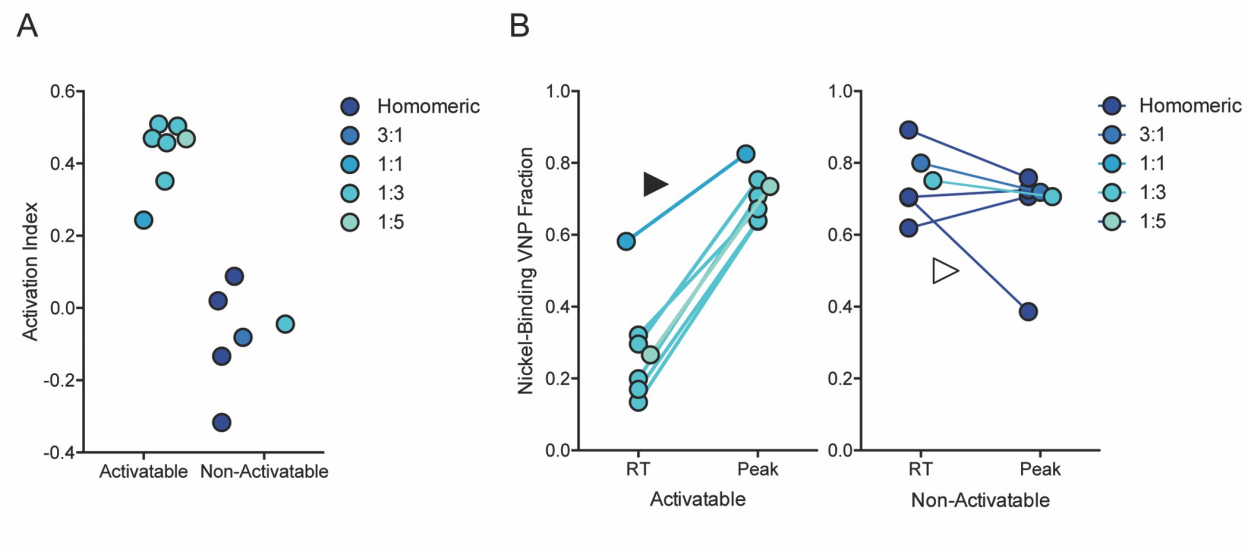



**Figure 2.6:** Summary of VNPs with activatable and non-activatable peptide display. A: Classification of VNPs as activatable and non-activatable. VNPs were grouped into activatable or non-activatable subsets using K-means clustering on the Activation Index into two clusters, resulting in an approximate threshold of 0.2. B: RT and peak binding (60–62 °C) of activatable and non-activatable VNPs. Activatable VNPs exhibit a shift from low to high binding between RT and peak, while non-activatable VNPs exhibit high binding at both temperatures. Black arrowhead indicates the $\Delta 30_1$-VP3$_1$ mosaic VNP, which exhibits higher RT binding than the other activatable VNPs. White arrowhead indicates $\Delta 20$ homomeric VNP, which is the only VNP to exhibit a >0.15 drop in binding at peak as compared to RT.

When we examine binding at RT *versus* binding at ~62 °C (the temperature of peak wt activatable peptide display) (Figure 2.6B), activatable viruses exhibit low binding at RT with the exception of $\Delta 30_1$-VP3$_1$. Binding typically increases at least two-fold ($\Delta 30_1$-VP3$_1$ excepting) from RT to peak temperature activation. Non-activatable viruses exhibit high column binding at both RT and peak temperature (Activation Indices generally less than 0.15), with the exception of the homomeric $\Delta 20$. This VNP exhibits a decrease in binding at peak temperature. The $\Delta 20$ does not show reduced genomic protection post incubation at 60 °C as compared to wt, indicating that these results are not solely due to capsid denaturation (Figure 2.7). Collectively, these data indicate that activatable viruses conceal N-terminal His$_6$ tags until stimulus detection, while non-activatable viruses have constantly exposed His$_6$ tags. Notably, the mosaic VNPs described in this work achieve higher functional output levels than the ΔPLA mutant we generated previously [172]. This improvement is apparent both in terms of maximum nickel column binding attained (83% for $\Delta 30_1$-VP3$_1$ as compared to 40% for ΔPLA) and Activation Index (>0.5 for $\Delta 10_1$-VP3$_3$ and $\Delta 20_1$-VP3$_3$ as compared to 0.2 for ΔPLA).

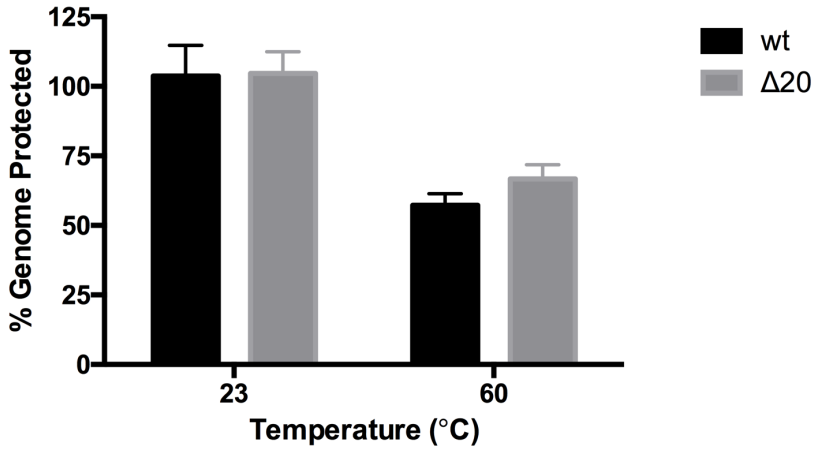



**Figure 2.7:** Benzonase genomic protection assay of Δ20 homomeric VNP post-incubation. Δ20 was incubated at the stated temperatures for 30 minutes, then benzonase genomic protection assay was conducted. wt AAV2 is included as a control. Differences between Δ20 and control are not significant. Error bars are SEM (N=3).

## 2.4 Conclusion

We have identified a key design parameter for engineering activatable peptide display in AAV-based nanodevices. The ratio of subunits in the mosaic capsid is the most influential design parameter for VNP activatability, where VNPs are most activatable when the subunit with the responsive peptide motif (*e.g.* VP2-$His_6$ truncation mutant subunit) is the minor capsid component. A potential structural explanation for this observation is that there is a limit to the number of longer-N termini that can be packed into the interior of the capsid, and the remaining N-termini are surface displayed prior to activation. Surprisingly, the length of the VP2-$His_6$ truncation mutants does not appear to be a significant factor in activatability, suggesting that the capsid subunit regions removed in the Δ10-Δ60 truncation mutants (corresponding to all but five of the VP2 residues located before the VP3 start codon) are not essential for activatable peptide display. Further investigation is required to determine why high proportions of VP2-$His_6$ truncation subunits result in incomplete N-termini packaging in the assembled capsid, and what regions (if any) on the capsid subunit N-terminus are required for activatable peptide display.

More broadly, the work described here presents an avenue for engineering activatable protein-based nanodevices by reprogramming naturally occurring viral capsid dynamics. This strategy may be applied to create nanodevices designed to carry out specific functions upon detection of extracellular or intracellular target stimuli. VNPs can incorporate a large number of linear cell-targeting or cell-penetrating peptides, facilitating specific tissue tropism and alternate paths to cytoplasmic entry [175,176]. Homomeric VNPs that are always "ON" may be used to display 60 copies of targeting peptides, while mosaic VNPs that exhibit activatable display may be further engineered to respond in a stimulus-responsive manner to biomarkers specific to particular cells or tissues. While the VNPs described here lack the VP1 unique domain required for endosomal escape and delivery of genetic material to the nucleus, they are internalized into cells similarly to wt (Figure 2.8), suggesting capsid-receptor interactions are retained. These VNPs may be used to deliver peptides to the endosomal environment, or to the cytoplasm with the assistance of endosomolytic peptides [177]. VNPs may also be targeted to the nucleus or other intracellular targets for combined gene and peptide

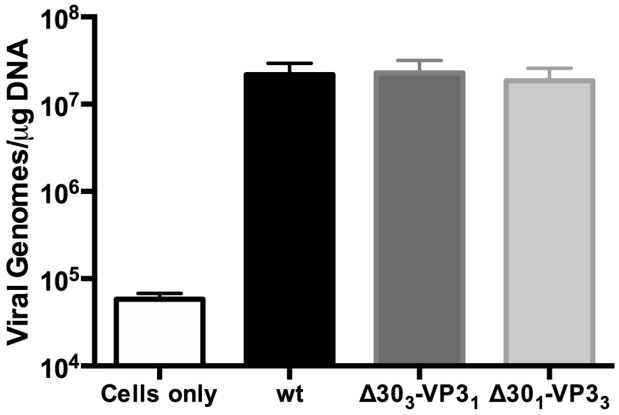



**Figure 2.8:** Cellular internalization of $\Delta 30_3$-$VP3_1$ and $\Delta 30_1$-$VP3_3$ VNPs. 1.8E6 confluent HEK293T cells were transduced with VNPs at 5,000 multiplicity of infection. After 2 hours, cells were washed with PBS 3 times and harvested. Intracellular DNA was extracted using E.Z.N.A. Tissue DNA Kit (Omega Bio-tek). Viral genomes were quantified with qPCR and normalized to total DNA extracted. wt AAV2 and cells without transduced virus are included as positive and negative controls, respectively. Error bars are SEM (N=2).

delivery using signal peptide sequences. Programming such peptide-based stimulus-responsive viral outputs may allow VNPs to interact with intracellular machinery to produce signal detection, protein function modulation, and targeting behaviors in an activatable or constant manner. When combined with viruses' innate ability to infect cells with high efficiency, this approach may lead to opportunities in VNP diagnostics and therapeutics.

## 2.5 Experimental

### 2.5.1 Cloning of virus variants

AAV2 mutants were created through site-directed mutagenesis of pXX2 plasmid using a PfuUltra High-Fidelity DNA Polymerase protocol (Stratagene). First, expression of all three proteins was prevented by silencing all start codons present on the *cap* gene, where the three VPs share an open reading frame [173]. To form viral truncation mutants, an ATG start codon followed by a $His_6$ tag-coding sequence was inserted into the VP2 N-terminus. Six VP2-$His_6$ mutants of decreasing capsid protein length were created by altering the location of insertion along regularly spaced, in-frame intervals. Mutant plasmids were sequence-verified *via* an external vendor (GENEWIZ).

### 2.5.2 Virus production and purification

To produce the AAV2 vectors, 7.5 $\mu$M polyethylenimine (PEI) was used to cotransfect the *rep-cap* encoding plasmid (pXX2 for wild-type AAV2 capsid, mutated plasmids for truncation variants), adenoviral helper plasmid pXX6-80, and pGFP (encoding GFP reporter gene flanked by ITRs) into HEK 293T cells, which acted as the site of protein production and viral assembly. Approximately 48 h after transfection, viruses were extracted by first lysing the cells with three freeze-thaw cycles. Excess and unpackaged nucleic acids were digested using 50 U/mL benzonase nuclease (Sigma) and removed with centrifugation. Fully-formed and packaged viruses were separated using iodixanol density gradient ultracentrifugation [178].

Viruses were applied to HiTrap Heparin HP columns (Amersham Biosiences) for further purification. The columns were washed with gradient buffer (10 mM Tris, 10 mM $MgCl_2$, 150 mM NaCl) and loaded with AAV in iodixanol. 5 mL elution buffer (10 mM Tris, 10 mM $MgCl_2$, 1 M NaCl) was applied to the heparin columns. Purified AAV was collected from the second elution fraction and dialyzed in 3 exchanges of 500 mL DPBS + Mg/Ca (55.9 mM $Na_2HPO_4$, 3.4 mM $MgCl_2{\cdot}6H_2O$, 18.5 mM KCl, 10.1 mM $KH_2PO_4$, 944.0 mM NaCl, 0.3 mM $CaCl_2{\cdot}2H_2O$) for a total of 18 h at 4 °C.

### 2.5.3 Virus titer determination

Quantitative polymerase chain reaction (qPCR) determined viral titer, or the effective concentration of viral genetic material. 2 M NaOH was used to break open viral capsids and release packaged DNA. Following neutralization with 2 M HCl and dilution with ultrapure water, SYBR Green PCR Master Mix (Life Technologies) and primers against the packaged CMV were added to the samples and run on the Bio-Rad CFX96 qPCR machine. Serially diluted rAAV plasmid DNA standards were also prepared and used to obtain absolute titer values.

### 2.5.4 Western blot of virus capsids

Western blot assays were used to analyze VP subunit composition in assembled AAV capsids. Viruses were first denatured using LDS sample buffer (4X) (Life Technologies). Samples were electrophoresed in 12% bis-tris gels according to manufacturer's protocol (Bio-Rad). The protein bands were wet-transferred to a nitrocellulose membrane (GE Healthcare), and the membranes were then blocked in 5% milk in PBS-T (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 2 mM $KH_2PO_4$, 0.2% Tween). The membrane was incubated first in monoclonal mouse B1 and $\alpha$-His antibodies (1:50 and 1:100 dilutions, respectively; American Research Products) followed by HRP-conjugated goat anti-mouse IgG antibody (1:2000 dilution; Jackson Immunoresearch). Lumi-Light Western Blotting Substrate solution (Roche) was added to the blot and the resulting stains were imaged using a GE Healthcare ImageQuant LAS 4000 imager. Densitometry analysis of gels was conducted using ImageJ 1.46r, as described in the ImageJ documentation (https://imagej.nih.gov/).

### 2.5.5 Transmission electron microscopy

300 mesh continuous carbon sample grids (Ted Pella) were glow discharged, and 8 $\mu$L of DPBS-purified virus was applied to the grid and left for 5 min, held in tweezers. The grid was then wicked dry with filter paper and washed twice by quick immersion in separate drops of 50 $\mu$L of ultrapure water, wicking dry in between each wash. Next, the samples were negatively stained by immersion in two drops of 50 $\mu$L uranyl formate (7.5 mg/mL filtered with 0.2 $\mu$m syringe filter; EMS), where the sample was left in the second drop for 20 s. The sample was wicked dry and left to air dry for 15 min. TEM images were taken with a JEM FasTEM 2010 transmission electron microscope.

### 2.5.6 Benzonase protection assay

DPBS-purified virus samples (2.5 $\mu$L) were diluted in 47.5 $\mu$L 1X endo buffer (1.5 mM $MgCl_2$, 0.5 mg/mL BSA, 50 mM Tris, pH 8.0) and mixed thoroughly. 20 $\mu$L of each sample was split into two separate tubes and incubated with 0.5 $\mu$L benzonase nuclease (250 U/$\mu$L, Sigma) or 0.5 $\mu$L sham buffer (50% glycerol, 50 mM Tris-HCl, 20 mM NaCl, 2 mM $MgCl_2$, pH 8.0) at 37 °C for 30 minutes. To terminate nuclease activity, 0.5 $\mu$L of 0.5 M EDTA was added to both. Viral titers for the benzonase-treated and sham-treated samples

were determined using qPCR, and genome protection was calculated by the ratio of benzonase-treated titer to sham-treated titer.

### 2.5.7 Nickel affinity columns

DPBS-purified virus samples were diluted to 1.5x$10^9$–3x$10^9$ viral genomes per column in 120–400 $\mu$L binding buffer, ensuring at most a 1:5 ratio of iodixanol to binding buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4). Samples were heated to the specified temperature for 30 minutes and applied to a His SpinTrap nickel spin column (GE Life Sciences) after equilibrating with 600 $\mu$L binding buffer. All spins were performed at 100 x g for 30 s. Samples were washed with 600 $\mu$L binding buffer and eluted with 200 $\mu$L elution buffer (20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4). Viral titers for the samples from the load, wash, and elution fraction were determined using qPCR and normalized to the total amount recovered.

### 2.5.8 Statistical analysis

Correlations between VNP characteristics were determined using Pearson correlation analysis in GraphPad Prism to compare data from all VNPs that formed at titers sufficient for nickel column analysis. Pearson correlations are computed as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Capsid assembly is defined as genomic titer normalized to a 10-plate prep (*i.e.* titers from one-plate preps are multiplied by 10). Genomic protection is defined as the percentage of genomes protected after benzonase treatment. Activation Index is defined as the difference in binding at the peak activation temperature and at RT.

# Chapter 3

# Developing an Adeno-Associated Virus-Based Protein Display Platform for Modular Delivery of Gene Editing Machinery

‡

Adeno-associated virus (AAV) has emerged as a promising vehicle for the delivery of gene-editing machinery, primarily due to its nuclear entry ability and low immunogenicity. However, AAV's genomic carrying capacity (4.7kB) is too small to package the well-characterized Streptococcus pyrogenes Cas9 (spCas9) gene along with standard promoters. AAV can fully package smaller Cas9 variants, but no room is left to spare for repair DNA templates. Additionally, AAV-delivered transgenes can persist episomally in cells for several years, resulting in prolonged Cas9 expression and increased off-target editing. To resolve these concerns, we have developed an AAV-based delivery system displaying Cas9 protein on the surface of the capsid. This delivery approach will result in transient presence of Cas9 in target cells limiting off-target editing, and free up space within the AAV capsid for repair template delivery and inclusion of reporter genes.

To attach Cas9 to the surface of AAV, we have functionalized the AAV capsid and Cas9 proteins with coiled-coil binding motifs. These motifs exhibit specific heterodimerization, allowing for AAV and Cas9 to be synthesized separately and avoiding the need for a combined purification protocol. We have established that both Cas9 and AAV retain their structural and functional properties with these coiled-coil motifs attached through gene editing activity and infectivity assays. Successful Cas9 attachment to AAV is demonstrated through column chromatography. The Cas9-AAV assemblies are currently being characterized to determine the copy number of Cas9 proteins bound per viral capsid. Cas9-AAV will be evaluated for infectivity and genome editing capacity *in vitro* and screened for immunogenicity through serum neutralization assays. Finally, this platform will be tested for *in vivo* genome editing in a mouse model. This coiled-coil display approach is modular and versatile in design, allowing for the display of multiple proteins in addition to Cas9. In the future, our engineered AAV-protein display platform may be used to create multifunctional

‡This chapter has been adapted from a publication in preparation: *Developing an Adeno-Associated Virus-Based Protein Display Platform for Modular Delivery of Gene Editing Machinery.* Nicole Thadani, Joanna Yang, Ciaran Lee, Buhle Moyo, Biki Kundu, Xue Gao, Gang Bao, and Junghae Suh

viral therapeutics capable of targeting tissues and delivering a variety of functional protein domains.

## 3.1 Introduction

Viruses are naturally-occurring nanodevices, evolved to sense and respond to environmental cues triggering delivery of their genomic cargo and self-replication [179]. Due to their innate strengths at protecting and delivering genetic material, viruses are an ideal gene therapy delivery platform. Adeno-associated virus (AAV) in particular has demonstrated potential as a gene therapy vector in clinical trials, due to its non-pathogenicity and low immunogenicity [4]. Synthetic virology has emerged as a strategy for optimizing AAV vectors, incorporating responses to novel environmental 'inputs' and delivery of non-native 'outputs' in the form of genes or biomolecules. Efforts have primarily focused on altering AAV's interactions with cellular receptors to improve targeting and modifying capsid epitopes to reduce immune recognition [180–187]. We have recently explored the AAV mediated delivery of non-genetic outputs to develop combination therapies [188, 189]. Here, we incorporate coiled-coil domains as non-covalent linkers in the AAV capsid as a tool for surface attachment of peptides and proteins.

Viral capsid protein attachment post-assembly has previously been attempted as a strategy for vector retargeting [190]. An initial approach used a bispecific antibody to bind both the AAV capsid and a cell-specific receptor found on human megakaryocytes [191]. Subsequent platforms conducted biotinylation of the AAV capsid for coupling with avidin-fusion proteins. Both non-specific biotinylation and specific metabolic biotinylation of small peptide insertions were used to attach a variety of targeting ligands and probes [192–194]. Biotinylation creates a modular capsid binding site capable of picomolar binding affinities, but attached proteins must be conjugated to streptavidin or avidin, large domains that may interfere with protein folding or function. Chemical conjugation approaches have also been applied to the AAV capsid, both non-specifically on surface-exposed lysines and at specific sites introduced using unnatural amino acids [188, 195–197]. These approaches have been used to attach targeting moieties and probes as well as chemotherapy drugs and nucleic acids. However, chemical conjugation reactions must be optimized to achieve high yield and frequently utilize catalysts that are not biocompatible, requiring additional purification steps [198, 199].

We developed a coiled-coil based binding platform for non-covalently linking proteins to the surface of the AAV capsid (Velcro-AAV). First postulated by Francis Crick and Linus Pauling, coiled coils are an ubiquitous biological motif, appearing in fibrous proteins, signal transduction cascades, membrane binding interfaces, and genomic DNA transcription factors [200, 201]. Coiled-coil motifs consist of a bundle of $\alpha$-helix domains wound into a periodic superhelix, with tight packing of the residue sidechains. The prototypical coiled coil consists of repeating heptads with hydrophobic domains stabilizing the superhelical structure and charged residues facilitating electrostatic interactions between $\alpha$-helices [202, 203].

Our platform uses the E3/K3 engineered leucine zipper pair, designed *de novo* by Litowski and Hodges based on principles gleaned from the study of naturally-occurring leucine zippers and optimized for maximum binding affinity, ease of purification, and shorter length [204–207]. This construct and similar leucine zippers have been widely adopted by protein engineers due to their small size and nanomolar binding affinity [208]. Their *de novo* design also reduces the potential for cross-talk with native signalling cascades. These zippers have the unique property of denaturing under endosomal pH, allowing for intracellular release of cargo [209].

Leucine zippers have previously been utilized for a diverse range of applications in protein engineering and delivery. Zippers have been used to attach fusion protein components after synthesis in cells and functionalized with cysteines to covalently link protein pairs [210,211]. Leucine zippers have also served as the linker between components of gold, protein, and polymer delivery vectors [212–217]. Some delivery systems

have capitalized on stimulus-responsive properties of leucine zippers to trigger nanoparticle aggregation or targeted drug release in response to environmental pH or temperature [218–221]. Researchers have also utilized leucine zippers to generate and interact with signal transduction components at the cell surface, triggering lysosomal uptake or aggregating transmembrane proteins to induce apoptosis [222–224]. Through these studies, researchers have demonstrated the modularity and biocompatibility of leucine zippers, along with their many advantageous properties as connectors in delivery vectors.

Here, we describe a surface-attachment system for AAV that uses a relatively small genetic insertion to facilitate attachment of potentially large proteins post-viral assembly. This approach is most useful for the modular attachment of large proteins that may disrupt capsid formation if inserted genetically, toxic proteins that may damage producer cells, or delicate proteins that are unlikely to retain activity when purified along with the AAV capsid. For example, antibodies and other large targeting moities may be attached to the capsid for modular retargeting to cell types without known small-peptide receptor ligands. Large proteinaceous cargo may also be tethered to the capsid surface, leaving the capsid interior space available for co-delivery of genetic material. Gene editing machinery in particular may benefit from this platform, as expression cassettes of the requisite proteins often exceed AAV's 4.7 kB genomic packaging limit, and the virus's genomic cargo space may be used to encode nucleic acid components of these systems.

CRISPR/Cas9 is a gene editing platform composed of a Cas9 nuclease guided by RNA that induces double-stranded breaks (DSBs) at precise locations in DNA [225]. Derived from a bacterial defense mechanism, the CRISPR/Cas9 system has found widespread use as a gene editing tool due to ease of retargeting through modification of the 20-bp guide RNA [226]. DSBs induced by CRISPR/Cas9 are then fixed through cellular DNA repair mechanisms. Non-homologous end joining (NHEJ) introduces insertion/deletion mutations disrupting gene function, while homology-directed repair (HDR) can be used to insert a desired gene or fragment from a co-delivered donor template [227]. CRISPR/Cas9 delivered as genetic material has treated neurological, muscular, and retinal diseases in animal studies [228–230]. However, persistent expression of Cas9 poses safety concerns due to the potential for off-target genome editing [227, 231]. An alternate approach is the delivery of Cas9 protein, either coupled to guide RNA as a ribonucleoprotein (RNP) or with guide RNA delivered separately. This strategy has been demonstrated to reduce off-target gene editing due to the transient presence of Cas9 protein before degradation [232]. Several vehicles have been explored for Cas9 protein delivery in cell culture and for local delivery *in vivo*. To overcome the challenge of cell entry, RNPs have been coupled with cell-penetrating peptides, enclosed in lipid coats, viral coats, or liposomes, tethered to gold nanoparticles, or delivered as naked RNPs through electroporation [232–237]. As a proof-of-concept, we are applying our Velcro-AAV platform to the delivery of Cas9 protein and guide RNA, both in RNP form and with guide RNA delivered in the viral transgene. This approach extends AAV's well-studied cell targeting and native nuclear entry properties to the Cas9 system, while avoiding potential concerns from the long-term expression of gene editing machinery delivered as a viral transgene.

## 3.2 Results

### 3.2.1 Design of Velcro-AAV VNPs

We developed a surface-attachment platform for AAV based on leucine zipper coiled-coil heterodimerization. The Velcro-AAV9 concept allows for modular attachment of protein components to the AAV capsid post-assembly (Figure 3.1A). To facilitate this, we designed genetic modifications to the AAV9 *cap* gene to incorporate a leucine zipper peptide motif after residue G453, a site previously demonstrated to tolerate insertion of small peptides and to result in surface-display [238]. We selected an engineered leucine zipper

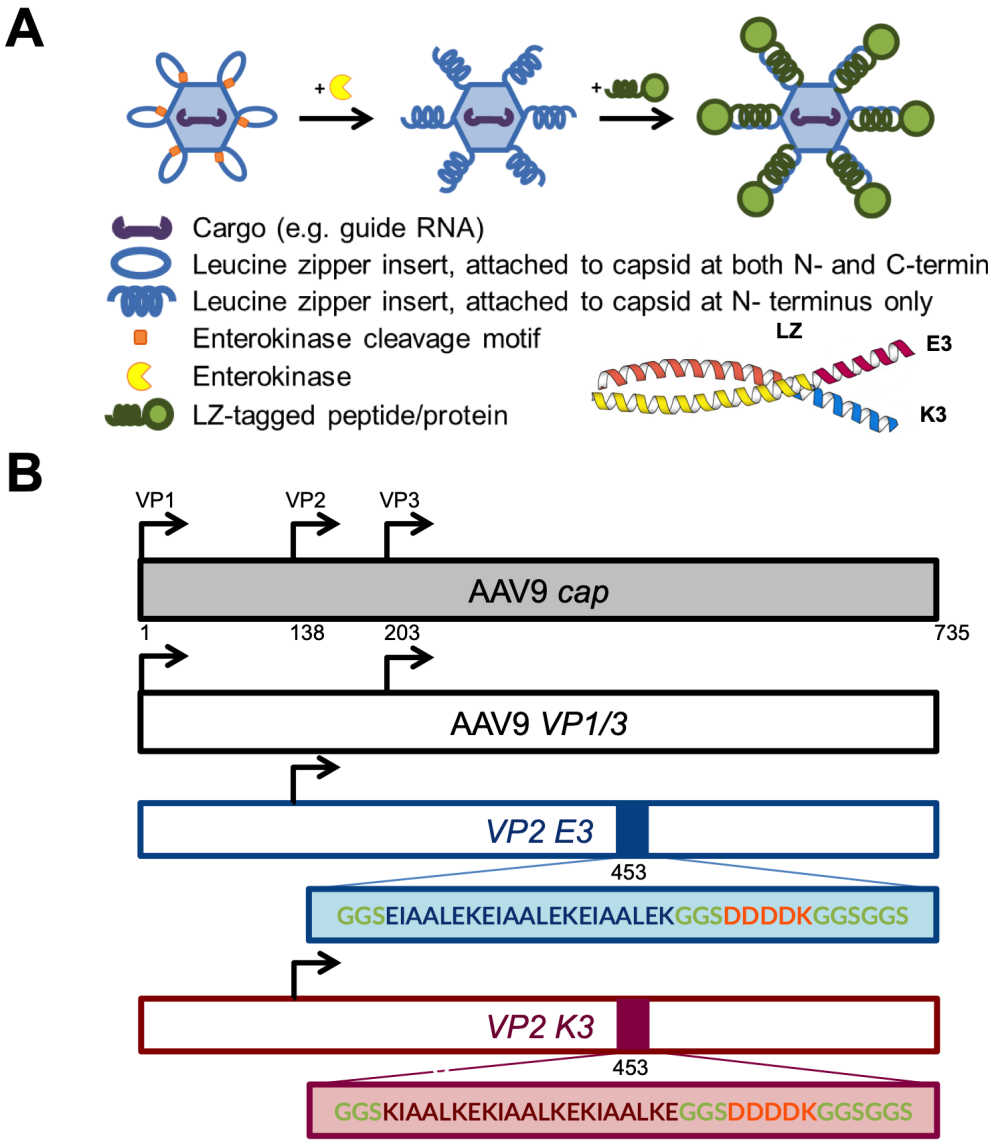



**Figure 3.1:** Design of Velcro-AAV platform. A) Schematic of Velcro-AAV system. The leucine zipper is inserted into the capsid in a loop conformation. Assembled viruses are enterokinase-digested to linearize the motif, making it available to bind to peptides and proteins incorporating a complementary leucine zipper heterodimer pair. B) Plasmids used to assemble Velcro-AAV. Leucine zipper sequences were inserted after residue G453 in the AAV9 *cap* gene, along with spacers (green) and an enterokinase recognition sequence (orange). Start codons for VP1 and VP3 were replaced with CTG. These modified E3 and K3 plasmids were co-transfected with AAV9 VP1/3 to assemble mosaic capsids.

motif with a short three-heptad (21 residues) sequence optimized for maximum stability [207] and designed peptide inserts for the K3 and E3 components of the heterodimer. As G453 occurs in the middle of the VP3 protein sequence, insertions at this point take on a loop conformation. To linearize the zipper, we incorporated an enterokinase digestion motif at the C-terminus of the zipper sequence. Enterokinase is an intestinal enzyme involved in digestion that has found use as a method for removing affinity tags from proteins post-purification [239, 240]. We also included GGS linkers to increase flexibility of the insertion,

facilitating VP folding and virus assembly [241, 242]. As capsid inserts have the potential to interfere with viral assembly and function, we used a mosaic capsid design to incorporate leucine zipper inserts on a fraction of capsid subunits. AAV typically assembles with the capsid subunits VP1, VP2 and VP3 incorporated in a 1:1:10 ratio [5]. In our design, inserts were incorporated in a modified AAV9 VP2-only plasmid. We then assembled Velcro-AAV through co-transfection of AAV9 VP1/3 and modified E3 VP2 or K3 VP2, with the aim of incorporating zippers on approximately 1/12th of capsid subunits (Figure 3.1B).

### 3.2.2 Characterization of Velcro-AAV9 VNP assembly, zipper incorporation, genomic protection, and transduction

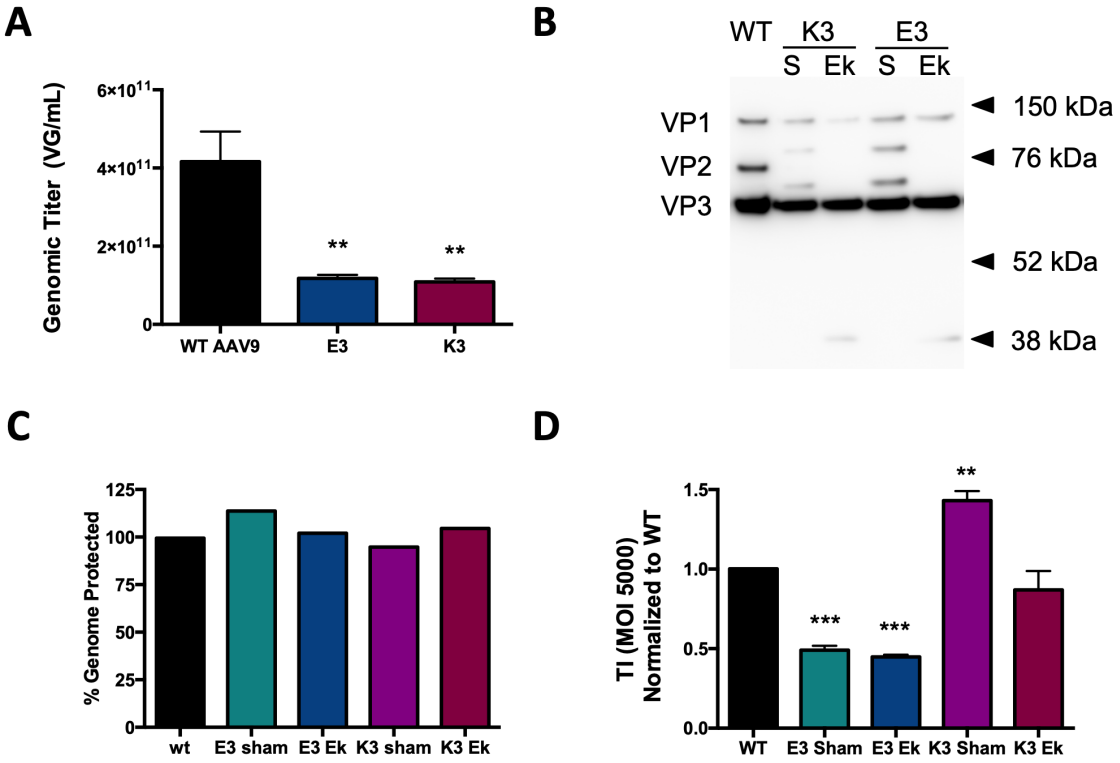



**Figure 3.2:** Characterization of Velcro-AAV formation, stability, and transduction. A) Genomic titers of 1-plate virus preps of E3 and K3 Velcro-AAV. Titers are expressed in terms of viral genomes per mL and were determined using qPCR. wtAAV9 is included as a control. N=3 independent virus preps, error bars SEM. One-way ANOVA was performed with Dunnett's post-hoc multiple comparison test to compare Velcro-AAV to WT. **p<0.01 B) Western blot of enterokinase and sham-treated samples. Purified viruses were treated with enterokinase or sham, denatured, and components analyzed on a western blot using B1 antibody. B1 antibody recognizes an epitope present in the C-terminus of all VPs. C) Benzonase genomic protection assay. Velcro-AAV samples pre- and post-enterokinase digestion were treated with either benzonase or sham buffer and the fraction of genomes remaining in benzonase-treated samples as compared to sham was quantified using qPCR. N=1 independent experiment (Repeats in progress). D) Transduction Assay. Cho-lec2 cells were transduced at an MOI of 5000 with Velcro-AAV. Transgene expression was measured by flow cytometry at 48h post-transduction. Transduction index (the product of %+GFP cells and geometric mean fluorescence intensity) is reported relative to wt. N=3 independent experiments, error bars SEM. One-way ANOVA was performed with Dunnett's post-hoc multiple comparison test to compare Velcro-AAV to WT. **p<0.01, ***p<0.001

Velcro-AAV capsids were assembled using 4-plasmid transfection, purified, and quantified. AAV9-E3 and AAV9-K3 form at 25% of wt AAV9 titers (Figure 3.2A). These reduced titers are comparable to the wt titers of other AAV serotypes with lower production yield than AAV9 [243]. Assembled capsids were digested with enterokinase or sham, resultant capsids denatured, and constituent proteins visualized on a western blot (Figure 3.2B). AAV9-E3 and AAV9-K3 both incorporate modified VP2-LZ proteins, as designed. Additionally, these viruses contain some modified VP3-LZ proteins, as indicated by the faint band directly above the VP3 band in sham-treated lanes. This is likely due to leaky transcription from the CTG codon used to replace VP3 start codons in the VP2-LZ plasmids, as previously observed with other genes in mammalian cells [244]. After digestion with enterokinase, the VP2-LZ and VP3-LZ bands disappear and the C-terminal cleavage product appears (predicted mass is 32 kDa), indicating complete digestion. Digested capsids were assayed for genomic protection against benzonase to determine if enterokinase treatment reduced capsid structural integrity (Figure 3.2C). In preliminary results AAV-E3 and AAV-K3 exhibit approximately 100% genomic protection, similar to wt, regardless of enterokinase treatment. Digested capsids were also screened for their ability to transduce cells (Figure 3.2D). Cho-lec2 cells were transduced with AAV-E3 and AAV-K3 treated with either enterokinase or sham buffer. AAV9-E3 exhibits a 50% drop in transduction index, while AAV9-K3's transduction index is not reduced as compared to wt. These results are cell-type specific, as AAV9-E3 post-enterokinase digestion exhibited comparable transduction to wt AAV9 in human umbilical vein embryonic cells (HUVECs) (Figure 3.3). While AAV9-E3's transduction is not impacted by enterokinase digestion, AAV9-K3 shows a slight reduction in transduction post-enterokinase digestion.

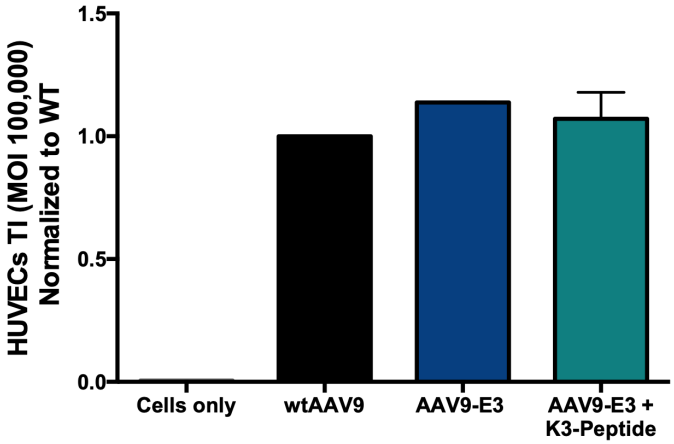



**Figure 3.3:** Velcro-AAV E3 transdution of HUVEC cells. Human umbilical vein embryonic cells were transduced at MOI 100,000 with wt AAV9, AAV9 E3 (Ek digested), and AAV9 E3 (Ek digested) + complementary K3 peptide. Transduction index is reported relative to wt. N=3 independent experiments, error bars SEM. One-way ANOVA was performed with Dunnett's post-hoc multiple comparison test to compare AAV9-E3 to WT - no significant differences were detected.

### 3.2.3 Velcro-AAV zipper binding

To determine if Velcro-AAV exhibit binding to their complementary zippers, we developed an assay based on nickel affinity chromatography. We designed leucine zipper peptides with C-terminal $His_6$ tags. AAV9-E3 was incubated with the K3-$His_6$ peptide and AAV9-K3 was incubated with the E3-$His_6$ peptide for 30 minutes at RT. Virus binding to a nickel affinity column was then quantified using qPCR.

We hypothesized that the Velcro-AAV would bind to complementary leucine zipper peptide, which would in turn bind to the affinity column through its $His_6$ tag. AAV9-E3, when incubated with 300,000x copies of K3-$His_6$ peptide, leads to almost 80% of capsids binding the column. However, a lower peptide concentration (3,000x) is not sufficient to induce virus column binding. At all peptide concentrations, wt AAV9 does not exhibit column binding, suggesting that this behavior is due to the interaction of the leucine zipper pair (Figure 3.4A). In this assay AAV9-K3 does not exhibit column binding — almost all virus is recovered in the flow through and wash fractions (Figure 3.4B). In subsequent studies we focused on AAV9-E3 due to these binding results.

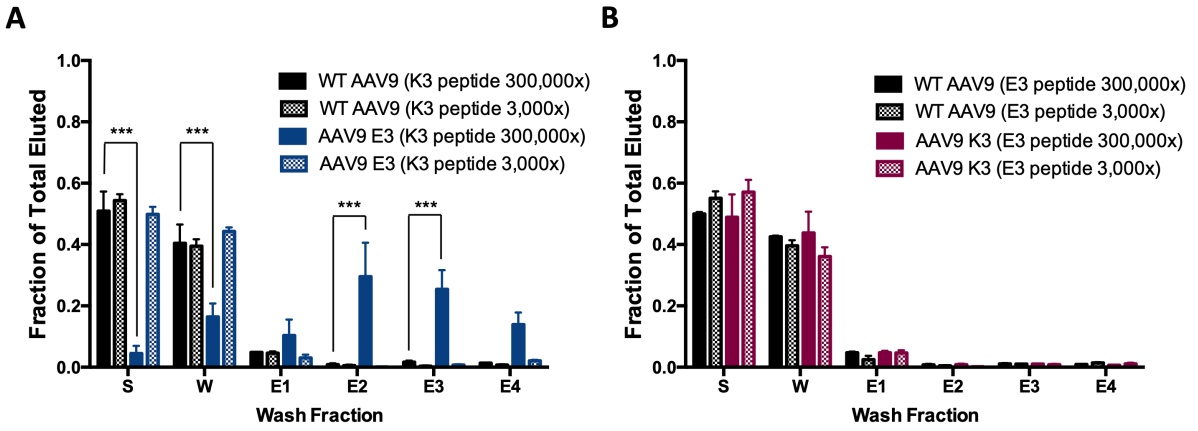



**Figure 3.4:** Characterization of Velcro-AAV interaction with complementary zipper peptides. Velcro-AAV were incubated with complementary $his_6$-tagged peptides (either 3,000x or 300,000x) for 30 minutes at RT, then the mixture was loaded on to nickel columns and the fraction of total virus eluted in the flow through, wash, and elution fractions was quantified via qPCR. N=3 independent experiments, error bars SEM. Two-way ANOVA was performed with Dunnett's post-hoc multiple comparison test to compare mutants to WT with 300,000x K3 peptide in each wash fraction. ***$p<0.001$

### 3.2.4 Design and functional characterization of zipper-modified Cas9

To design a zipper-tagged spCas9, we incorporated leucine zippers at either the N- or C-terminus of the *cas9* gene. We transfected HEK293T cells with these variants and guide RNA, then assayed for genome editing using the T7E1 assay. T7E1 endonuclease cleaves at mismatches in dsDNA, so mismatches introduced by NHEJ to repair Cas9-induced DSBs can be detected through gel electrophoresis. All variants exhibited genome editing comparable to wt Cas9, although the N-terminal K3 variant shows a slight reduction in indel frequency (Figure 3.5). The N-terminal K3 variant was selected for further analysis, as AAV-E3 displays the E3 zipper with a free C-terminus and the E3 and K3 zipper components bind in parallel, so K3N-Cas9 would theoretically bind to AAV-E3 with minimal steric hindrance [209].

To determine K3N-Cas9 affinity for AAV-E3, we used a nickel affinity chromatography assay. AAV-E3 was incubated with K3N-Cas9 with a C-terminal $His_6$ tag at 37 °C for 30 minutes. The mixture was then loaded on to a nickel column, and bound virus quantified using qPCR. 60% of AAV9-E3 is bound to the column when incubated with 300,000x K3N-Cas9, indicating that K3N-Cas9 and AAV-E3 exhibit binding (Figure 3.6). This interaction is not detected with the wt AAV9 virus, suggesting that this interaction is zipper-mediated. Lower concentrations of K3N-Cas9 protein do not result in AAV9-E3 column binding.

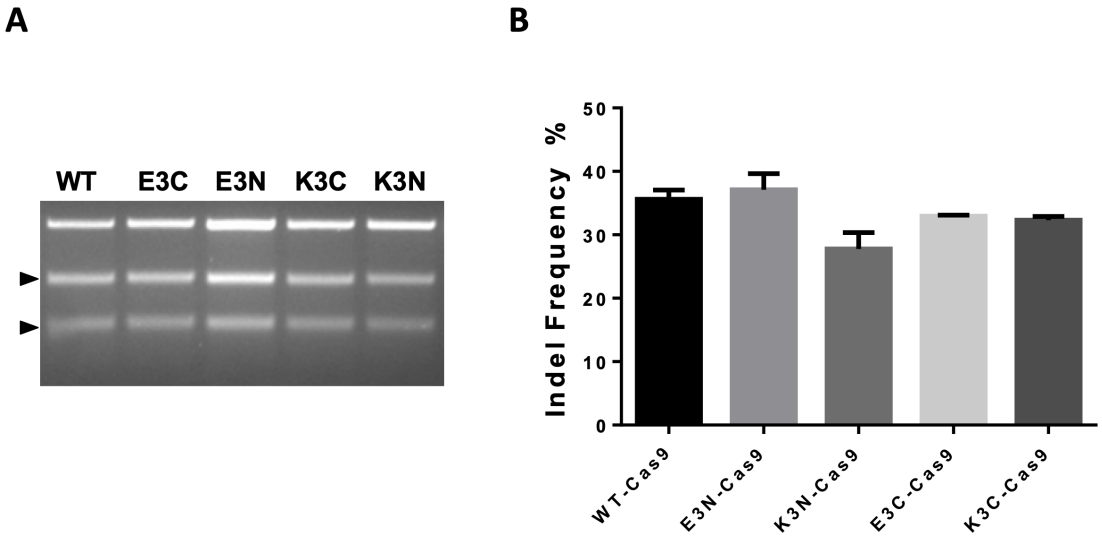



**Figure 3.5:** Leucine-zipper tagged spCas9 activity. Leucine zippers were inserted at the N- or C-terminus of Cas9. LZ-Cas9 were assayed for editing capacity by the quantification of insertions and deletions using a T7E1 assay. A) T7E1-digested DNA. Fragments (arrows) indicate genome editing at the targeted site. B) Quantification of % gene modification using ImageJ. N=2 independent experiments, error bars SEM.

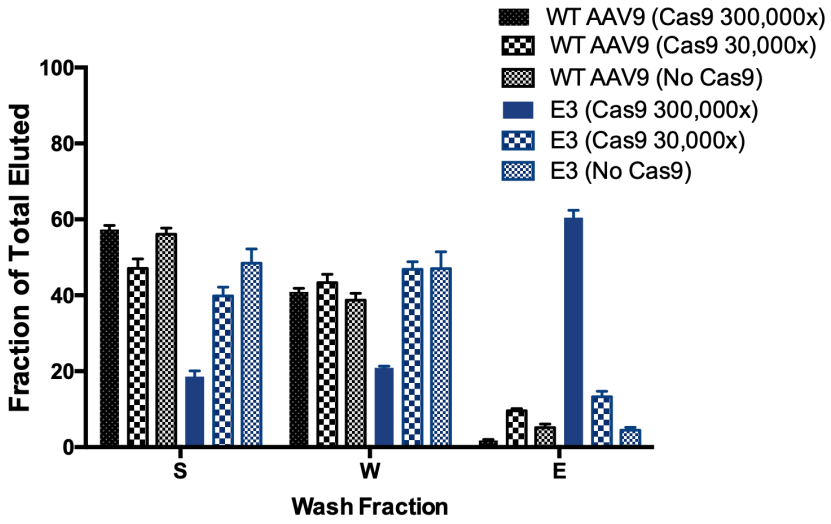



**Figure 3.6:** Quantifying the interaction between leucine-zipper tagged spCas9 and Velcro-AAV. AAV9-E3 was incubated with either 300,000 or 30,000 K3N-Cas9 copies per virus, then loaded onto a nickel column and the sample flow-through, wash, and elution fractions quantified. The elution data shown is the sum of 3 elutions with 250mM, 500mM and 1M imidazole. N= 2 independent experiments, error bars SEM.

## 3.3 Discussion

To facilitate modular attachment of retargeting motifs and protein cargo to AAV, we integrated one half of a leucine zipper heterodimer pair into the AAV9 capsid surface. Velcro-AAV capsids show a 3.5-fold reduction in vector production although many other inserts at the same site have not impacted capsid assembly and genome packaging [238]. Only one of the inserts previously tested at this site formed at

reduced titers, approximately 25% of wt AAV9 [245]. Tolerance of the AAV capsid to insertion has been demonstrated to vary with the size and sequence of insertion [46, 173, 246, 247]. The leucine zipper insert is 38AA long, while the longest insert previously tested at this site is 32AA long. The AAV9-E3 insert has a total charge of -6 and the AAV9-K3 insert has a neutral total charge, while previous inserts with no impact on genomic titer have been negatively charged (-2 to -8) [245]. These reduced vector titers may pose challenges in production scale-up. To mitigate this issue, the plasmid ratio of VP2-LZ and VP1/3 used in vector production could be adjusted to reduce the number of zipper-containing subunits produced and available for capsid incorporation. Alternatively, the VP2-LZ plasmid could be redesigned with a strong start codon at VP2, which has previously been demonstrated to reduce VP3 expression [173], thus blocking the leaky expression of VP3-LZ observed from this plasmid and reducing the total zipper-containing subunits formed.

Velcro-AAV require an enzymatic digestion step to linearize the leucine zipper motif before attaching protein cargo. This digestion step does not appear to impact capsid integrity, as Velcro-AAV post-enterokinase digestion exhibit genomic protection similar to wt AAV9. As this platform is further developed for *in vivo* applications, it will be useful to determine if this enterokinase cleavage step has any impact on capsid stability in response to varied pH or temperature conditions that may result in challenges for Velcro-AAV production pipelines or long-term storage [248]. The two Velcro-AAV designs have different transduction capacities. AAV9-E3 shows a reduction in transduction, while AAV9-K3 transduces cells similarly to wt AAV9. This may be a charge-based effect, as the AAV9-E3 insert has an overall negative charge while the AAV9-K3 insert has a neutral charge. Negatively charged domains have previously been used to interfere with AAV9 galactose receptor binding at this insertion site [238, 245]. Future designs of Velcro-AAV could mitigate this effect by incorporating positive charges to neutralize the impact of the E3 zipper domain. Alternatively, leucine zippers could be repositioned at the N-terminus of VP2, thus removing the need for enterokinase digestion and insertion of the negatively charged enterokinase cleavage domain [173]. Insertions at this site would also likely present on the capsid surface at a different location than the galactose-binding domain.

The AAV9 E3 construct is able to bind a K3-$His_6$ peptide as observed in a nickel affinity chromatography-based assay, but the AAV9 K3 construct does not show evidence of binding to an E3-$His_6$ peptide. The AAV9 K3 insertion may interact with the surrounding capsid, limiting its capacity for interaction with the E3 peptide. It is also possible that the E3-$His_6$ peptide takes on a conformation that prevents interaction with AAV9-K3 or with the nickel affinity column, potentially due to interactions between the positively charged $His_6$ domain and the negatively charged E3 domain. Although we observed AAV9 E3 binding to both the K3-$His_6$ peptide and K3N-Cas9 protein via the nickel affinity assay, 300,000x more peptide/protein than virus was required to observe this interaction. These results may be due to the tendency of K3 peptide and K3-fusion protein constructs to homo-oligomerize at high concentrations [209, 210]. We hope to optimize peptide concentrations during binding as well as binding temperature and buffer to reduce the amount of peptide/protein required.

## 3.4 Conclusion and future directions

To characterize the interaction of AAV9 E3 and K3N-Cas9, we plan to purify bound particles using nickel affinity chromatography to remove free AAV9 E3, followed by centrifugal concentration with a 1,000 kDa filter to remove free K3N-Cas9. The nickel affinity chromatography assay previously described established that this approach will remove free AAV9 E3. To confirm that centrifugal concentration removes free K3N-Cas9, we will conduct western blots to monitor free K3N-Cas9 in the flow-through. Purified bound particles

will be titered and analyzed using a quantitative ELISA for Cas9 to determine the number of K3N-Cas9 bound to each AAV9 E3 capsid. Bound particles will also be assayed for infectivity through transduction assays in Cho-Lec2 cells.

The utility of Velcro-AAV/Cas9 particles as a gene-editing platform will be determined by transduction of HEK293 cells stably expressing GFP. We will quantify Cas9 activity both through the reduction of GFP expressing cells quantified by flow cytometery, and the rate of insertions and deletions quantified by T7E1 assay. Additionally, we will screen a RNP platform where K3N-Cas9 is complexed with GFP guide RNA before attachment to AAV9-E3. This approach removes the step of GFP guide transcription from the AAV transgene, increasing the likelihood of successful editing.

The Velcro-AAV platform expands AAV as a delivery platform for proteins, RNPs, and peptides in addition to genes. This approach may be used to develop combination therapies using both protein and genetic material to treat challenging diseases. Additionally, Velcro-AAV can be functionalized with reporters or with receptor-binding motifs, increasing the available toolkit for retargeting and monitoring gene therapy vectors. Extensions of this approach may draw on prior research developing stimulus-responsive coiled-coil motifs to promote the release of zipper-attached cargo at disease sites, or using cross-linking to strengthen coiled-coil linkers for systemic *in vivo* delivery. Gene editing machinery is an attractive cargo for Velcro-AAV, as this platform frees up vector cargo space to package template material for homology-directed repair and avoids problems with drawn-out expression of gene editing machinery that may lead to off-target effects. Our Velcro-AAV/Cas9 system may lead to the development of a safer and more efficient *in vivo* gene editing approach.

## 3.5 Experimental

### 3.5.1 Cloning of Velcro-AAV constructs

AAV9 VP2 LZ constructs were created through modification of the pXR9 plasmid. NgoMIV and KasI restriction digest sites were inserted after residue 453 in the AAV9 *cap* gene and extraneous restriction sites removed from the backbone through site directed mutagenesis (SDM) using the QuikChange protocol with Pfu Ultra polymerase (Agilent Technologies). To eliminate template DNA reaction mixes were digested with DpnI (New England Biolabs). Similarly, start codons for VP1 and VP3 were removed using SDM.

K3 and E3 zipper inserts with enterokinase cleavage sites, spacers, and NgoMIV and KasI restriction digest sites were designed and purchased as gBlocks (Integrated DNA Technologies). Backbone plasmid and gBlocks were double-digested using NgoMIV and KasI (New England Biolabs) according to manufacturer protocol. Reactions were heat-inactivated. The backbone plasmid was dephosphorolyated using antarctic phosphatase (New England Biolabs) and gel purified using a Zymoclean gel DNA recovery kit (Zymo Research). Inserts were purified using a DNA clean and concetrator kit (Zymo Research). Backbone and insert fragments were ligated using T4 DNA ligase (New England Biolabs) according to manufacturer protocol. Plasmid products were sequence-verified through an external vendor (Genewiz).

### 3.5.2 Virus production and purification

Velcro-AAV were created by a quadruple-plasmid polyethylenimine transfection in HEK 293T cells consisting of 1) pXX6-80 helper plasmid encoding adenoviral helper genes, 2) pITR-iRFP(GFPguide) encoding an RFP transgene and GFP guide RNA flanked by inverted terminal repeats, 3) VP1/3 (pXR9 with VP2 start codon knocked out) and 4) VP2 LZ construct. After 48 hours, cells were lysed through 3 freeze-thaws, then

treated with 50 units/mL benzonase nuclease (Sigma) and centrifuged to remove excess nucleic acids. The supernatant was layered atop a 15%-54% iodixanol step gradient, and the tubes centrifuged at 48,000 rpm in a Beckman Type 70Ti rotor for 105 minutes at 18 °C. Viruses were then extracted from the 40% iodixanol layer and concentrated into gradient buffer (10 mM Tris, pH 7.6, 10 mM $MgCl_2$, 150 mM NaCl) + 0.001% pluronic F-68 (GB-PF68, Thermo Fisher) using Amicon Ultra 100 kDa centrifugal filters (EMD Millipore).

### 3.5.3 Virus quantification

Viral titers were quantified using quantitative polymerase chain reaction (qPCR). Viral capsids were denatured in 2M NaOH at 75 °C for 30 minutes to release genomes. The mixture was then neutralized using 2M HCl. Samples were diluted in ultrapure water and reactions prepared using SYBR Green PCR Master Mix (Life Technologies) and primers targeting the CMV promoter. Serially diluted recombinant AAV transgene was used to generate a standard curve. Samples were analyzed using a BioRad CFX96 qPCR machine and absolute titers were calculated based on the standard curve.

### 3.5.4 Enterokinase digestion

Leucine zipper motifs in the AAV capsid were linerized through digestion with enterokinase. Virus was incubated with either sham (20 mM Tris-HCl, 200 mM NaCl, 2 mM $CaCl_2$, 50% glycerol, pH 7.2) or enterokinase enzyme (Light Chain, NEB) in 2 mM $CaCl_2$ at 23 °C for 24 hours. To remove enterokinase from the virus digest, Pierce microcentrifuge spin columns (Thermo Fisher) were prepared with Trypsin-inhibitor agarose beads (Sigma). Beads were prepared by flowing stripping buffer (0.1 M NaCl, 0.1 M Formic acid) through the columns three times, followed by 1x GB-PF68. The virus-enterokinase reaction was incubated with the Trypsin-inhibitor agarose beads on the column for 15 minutes with agitation, before spinning at 2000 xg for 2 minutes to elute virus.

### 3.5.5 Western blot of virus capsid proteins

Viral capsid proteins were characterized using western blot to determine capsid composition and monitor enterokinase cleavage. Viruses were denatured with LDS sample buffer (NuPAGE, Life Technologies). Samples were electrophoresed in 4-12% Bis-Tris gels (NuPAGE, Life Technologies) and bands wet-transferred to a nitrocellulose membrane. Membranes were then blocked in 5% milk in PBS-T (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 2 mM $KH_2PO_4$, 0.2% Tween) and washed in PBS-T. Membranes were then incubated in monoclonal mouse B1 primary antibody (1:50 dilution, American Research Products). Membranes were washed, then incubated in HRP-conjugated goat anti-mouse IgG secondary antibody (1:2000 dilution, Santa Cruz Biotechnology). Blots were then PBS-T washed and treated with LumiLight Western blotting substrate (Roche Applied Science). Blots were imaged using a GE Healthcare ImageQuant LAS 4000 imager.

### 3.5.6 Benzonase genomic protection assay

Virus samples treated with sham or enterokinase were diluted 1/10 in endo buffer (1.5 mM $MgCl_2$, 0.5 mg/mL BSA, 50 mM Tris, pH 8.0). Samples were divided into two 20 $\mu$L reactions and either 0.5 $\mu$L benzonase nuclease (Sigma) or 0.5 $\mu$L sham buffer (50% glycerol, 50 mM Tris-HCl, 20 mM NaCl, 2 mM $MgCl_2$, pH 8.0) was added. Reactions were incubated at 37 °C for 30 minutes, then nuclease activity was terminated through the addition of 0.0125 $\mu$M EDTA. Remaining genomes were quantified though qPCR

and genomic protection was calculated as the ratio of genomes in benzonase-treated samples to genomes in sham-treated samples.

### 3.5.7 Quantification of virus transduction

Viruses pre- and post-enterokinase cleavage were evaluated for their transduction abilities. Cho-Lec2 cells were seeded on tissue-culture treated poly-l-lysine 48-well plates. At 70% confluency, cells were transduced with virus in serum-free media at 5000 multiplicity of infection (MOI). Media was changed to media with serum at 24 h after transduction. Cells were harvested at 48 h post transduction and analyzed through flow cytometery (BD FacsCantoII). The transduction index (TI), a linear indicator of transduction efficiency, was computed as the product of %GFP+ cells and the geometric mean fluorescence intensity (gMFI) [147].

### 3.5.8 Nickel column zipper binding assay

3E9 vg of amicon-purified Velcro-AAV were diluted in 120 $\mu$l binding buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4) along with his-tagged complementary zipper peptide (purchased from GenScript) or his-tagged Cas9 with complementary zipper. Samples were incubated at RT (peptides) or 37 °C (K3N-Cas9) for 30 minutes. His SpinTrap columns (GE Life Sciences) were equilibrated with binding buffer. All spins were conducted at 100 xg for 30 s. Samples were loaded on to the columns, then columns washed with 600 $\mu$L binding buffer. Samples were eluted with four elution buffers with increasing concentrations of imidazole (20 mM sodium phosphate, 500 mM NaCl, [100, 250, 500, 1000] mM imidazole, pH 7.4). Each flow through was collected and viral titers determined using qPCR. Titers were normalized to the total amount of virus recovered from the load, wash, and elution fractions.

### 3.5.9 Cas9 activity quantification

Modified zipper-Cas9 activity was assayed via plasmid transfection of zipper-Cas9 genes and guide RNA using lipofectamine 2000 (Thermo-Fisher) into U2OS-EGFP HEK293T cells. Genomic DNA was harvested using QuickExtract DNA extraction solution (Epicenter). PCR was conducted to amplify genomic loci and amplified fragments cleaned using AMPure XP magnetic bead purification system (Agencourt). Amplicons were denatured and reannealed, then digested with T7 endonuclease I as previously described [249]. Gels were quantified using ImageJ and the area of peaks representing Cas9-induced cleavage fragments quantified as a percentage of peaks representing total DNA in the undigested amplicons. This fraction was used to calculate gene modification as previously defined (Equation 3.1) [250].

$$\% \text{ gene modification} = 100 \times (1 - (1 - \text{fraction cleaved})^{1/2}) \tag{3.1}$$

# Chapter 4

# Evaluation of Modeling Approaches to Predict Formation and Function of Genetically Modified Adeno-Associated Virus Capsids

§

Adeno-associated virus (AAV) is a promising gene therapy vector due to its efficient gene delivery properties coupled with relatively mild immunogenicity. To improve vector tissue- and cell-specificity and delivery efficiency, researchers are implementing a variety of combinatorial and rational design-based strategies for generating novel AAV-based vector libraries. Such mutational strategies frequently create large populations of non-forming or non-infective variants that reduce the effective depth of these libraries, costing time and resources through the screening process. We evaluated two different computational techniques that predict the fitness of proteins for their effectiveness at screening AAV mutants *in silico*. These approaches may facilitate the development of AAV capsid libraries with higher proportions of functional variants, accelerating screening and identification of promising therapeutic candidates. The first computational tool, called the frustratometer, uses thermodynamic modeling of protein structures to identify key regions facilitating protein-protein interaction and structural transformation. We applied this approach to AAV2 and identified candidate residues favoring assembled and disassembled states of the capsid. The second computational strategy, called direct-coupling analysis (DCA), is a statistical framework based on the principle of residue coevolution within proteins. DCA was used to make an additional set of predictions about the functionality of these capsid mutations. We then experimentally quantified the impact of these mutations on virus formation and transduction ability. Notably, the frustratometer-based metric shows some correlation with virus formation, but a metric based on the collection of DCA parameters is highly correlated with virus transduction ability, suggesting that global models of amino acid connectivity could be useful in predicting AAV function *in silico*. Our results suggest that structure-based and coevolutionary computational models

§This chapter has been adapted from a publication in preparation: *Evaluation of Frustration and Direct Coupling Analysis Computational Models to Predict Formation and Function of Genetically Modified Adeno-Associated Virus Capsids.* Nicole Thadani, Kiara Reyes Gamas, Qin Zhou, Susan Butler, Nicholas Schafer, Faruck Morcos, Peter Wolynes, and Junghae Suh

may be able to help elucidate the complex residue-residue interaction networks that are essential for virus assembly and function — information that could accelerate the process of viral vector design in the future.

## 4.1 Introduction

Adeno-associated virus is widely favored as a gene therapy vector, currently tested in over 200 human clinical trials internationally. AAV was recently approved by the US FDA as the first gene replacement therapy for an inherited genetic disorder, in part due to the vector's non-pathogenicity and generally benign safety profile [2, 3, 5, 251]. AAV is an efficient delivery vector, capable of transducing both dividing and non-dividing cells to produce sustained gene expression [252–255]. However, further improvements to the vector are highly sought to enhance delivery efficiency in targeted cell populations, which would minimize vector doses needed to achieve therapeutic effect and reduce production costs. Lower doses would also reduce any dose-dependent immune responses against the vector itself, mitigating a potential safety issue in high-dose applications [256–259].

To obtain optimal vector performance for a variety of disease targets, gene therapy developers may select from the hundreds of AAV variants discovered through biomining efforts. Researchers have employed a variety of design strategies to further engineer these AAV vectors for enhanced targeted delivery. Three main approaches that have emerged are 1) rational design, 2) directed evolution, and 3) bioinformatics-driven design. Rational design approaches may integrate protein components drawn from other biological sources into the viral capsid to endow new capabilities upon the virus [15]. Directed evolution explores the mutational space of the viral capsid by conducting rounds of mutation with tailored selection schemes to achieve a new phenotype [186]. Bioinformatics-driven strategies have recently emerged as a method drawing upon large datasets of AAV capsid information and applying computational models to accelerate exploration of the viral fitness landscape to find functional variants with desired properties. Initial studies in this space have yielded promising results, including AAV chimera populations designed using the SCHEMA algorithm to minimize structural disruption, mapping of AAV capsid amino acids important for structure and function using high-throughput analysis of large mutant libraries, and the development of ancestral AAV vectors [260–264]. Here, we explore a series of data-driven computational approaches with varying degrees of complexity for predicting the fitness of AAV capsid mutants with the future goal of incorporating these *in silico* approaches into the AAV vector design pipeline.

Virus intracapsid interactions are governed by the same thermodynamic principles at play in all protein-protein interactions. Protein-protein interaction is typically governed by weak non-covalent interactions, such as hydrogen bonding, salt bridges, van der Waals forces, and hydrophobic interactions [265]. The predominant interaction in capsid assembly is burial of hydrophobic regions between capsid monomer binding interfaces, which may be modeled as a mass-balance equilibrium [266, 267]. We sought to capture these fundamental capsid interactions through a coarse-grained molecular dynamics approach for simulating protein dynamics.

The frustration model is based on the energy landscape theory of protein dynamics [268]. For a protein to fold robustly, rapidly and maintain a single native structure, it is necessary for the energy landscape to take on a smooth, funnel-like shape, minimizing kinetic traps (Figure 4.1). This shape arises through evolutionary optimization, reducing the number of favorable potential residue-residue contacts in the protein, or "frustration" [269, 270]. This theory of minimal frustration does not preclude proteins from having some degree of residual frustration, manifesting in non-zero entropy with multiple valleys at the base of the energy landscape funnel [271]. This residual frustration may result from adaptations to facilitate protein allostery

and multimeric interactions. Frustration can be computationally quantified using the frustratometer, which draws upon coarse-grained modeling to simulate protein energy landscapes [272].

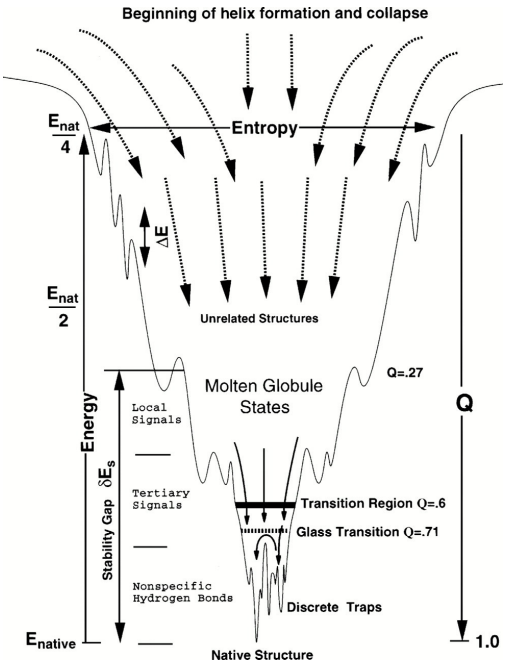



**Figure 4.1:** Protein folding energy landscape funnel. E is the solvent-mediated energy, and Q represents the fraction of native contacts. Hydrogen bonds, tertiary signals and local signals contribute to the stability gap, contributing to specific folding into a native structure. Reprinted with permission from [273].

Previous studies have shown that analyzing frustration patterns can provide insight on protein dynamics and interactions. Frustration analysis of proteins that demonstrate allostery indicate known regions of conformational shifting such as hinges are highly frustrated [270]. The frustration model has also been verified as predictive in determining homodimer and heterodimer docking sites and intermediate state interactions [274]. This model has also been applied to explore reaction mechanisms and intermediate conformations in a variety of protein behaviors. Transition states in the allosteric pathway of I$\kappa$B mediated disruption of NF-$\kappa$B DNA binding have been postulated using frustration analysis [275]. Here, we apply the frustratometer to capture interactions required for AAV capsid assembly and stability and predict destabilizing mutations. Studies drawing upon coarse-grained molecular dynamics modeling have previously captured viral assembly pathways, suggesting that this approach may prove fruitful in understanding this crucial component of engineered virus functionality. [276–280].

While thermodynamic interactions driving AAV capsid assembly and stability are crucial to viral function, AAV capsid proteins also interact with other viral components, helper-virus assembly proteins, and host cellular factors to successfully form genome-packaging viral particles and transduce cells [6, 32, 34, 41]. These necessary capsid interactions, in addition to the energy landscape required for AAV capsid protein folding and assembly, manifest as evolutionary constraints on the capsid proteins. For viral proteins these constraints are particularly stringent, as the small size of viral genomes drives the development of multifunctionality [281, 282]. These constraints result in pairs of capsid residues, such as residues in contact in the virus's assembled structure or residues that together interact with a cellular binding partner, exhibiting correlation across sequences in a protein family [283, 284]. Such coevolution of protein residue pairs may be quantified through statistical models to develop a map of the network of protein interactions underlying

structure and function [285].

Coevolution modeling seeks to identify residues with structural or functional relationships within a protein through analysis of the protein sequence family. Direct coupling analysis (DCA) has emerged as a prominent statistical framework for identifying direct physical interactions of protein residues from genomic data [286–288]. DCA is a global maximum entropy model that infers the parameters of an energy function with terms representing single-residue identity and pairwise couplings. By generating a global model of sequence families rather than a local model of each residue-residue interaction, DCA can isolate true interacting pairs from secondary correlations that appear due to the intertwined networks of primary interactions. When applied to a family of protein sequences, this model can produce ranked residue contact pairs as well as the probability that a given sequence is a member of the sequence family [289]. DCA has accurately recapitulated known structural contacts in a variety of proteins and intermediate protein structures [286, 287, 290].

DCA approaches were initially applied to the characterization of bacterial two-component system signalling networks, and the design of libraries predicted to enhance signalling [291–293]. More recently, DCA approaches have shown promise in identifying interacting residues in protein oligomers [294, 295]. Coevolution approaches previously applied to viruses have correctly identified interactions between viral proteins, but these models have not been evaluated for their ability to predict mutant virus functionality [296–298]. Here, we evaluate DCA's ability to capture patterns that identify functional members of the parvovirus family by predicting the formation and function of AAV mutants.

In this study, we applied the frustration model to AAV2 monomer and multimer assembly structures to identify residues with large shifts in frustration index between these states. We then mutated these residues to alanine and experimentally evaluated virus formation, thermal stability, and transduction. We compared these experimental results to predictions from the frustration and DCA models – the frustration approach shows promise for predicting aspects of mutant capsid structure and stability, while DCA may be useful for predicting capsid transduction.

## 4.2 Results

### 4.2.1 Frustratometer analysis of AAV capsid

We applied the frustratometer to AAV2 to predict residues that favor either the assembled capsid structure or the monomeric capsid protein. The AAV capsid consists of three capsid proteins, VP1, VP2, and VP3, that assemble in a 1:1:10 ratio to form a 60-mer capsid. Cryo-EM resolved structures of AAV2 do not include the unstructured VP1 and VP2 N-terminal domains, so this analysis focuses on the VP3 domain shared by all three VPs that forms the exterior capsid structure. As the 60-mer capsid structure is computationally intensive to analyze with coarse-grained modeling, a multimer assembly substructure was generated from the 60-mer capsid structure by selecting the eight VP3 subunits within 4.5 Å of a central monomer (Fig. 4.2A). Then, mutational frustration was calculated for every possible residue interaction in the monomer and multimer (Fig. 4.2B-C). The multimer structure contains more minimally frustrated residue pairs than the monomer, particularly at capsid subunit-subunit binding interfaces. This corresponds with previously observed shifts in frustration at protein binding interfaces related to the burial of hydrophobic protein domains [299].

Single-residue mutational frustration indices (looking at the frustration of each residue, as opposed to each residue pair) were also computed for the monomer and the multimer assembly. The AAV2 monomer has 30.6% minimally frustrated residues and 8.1% highly frustrated residues, while the AAV2 multimer has 30.3% minimally frustrated residues and 6.6% highly frustrated residues. Both capsid subunit configurations

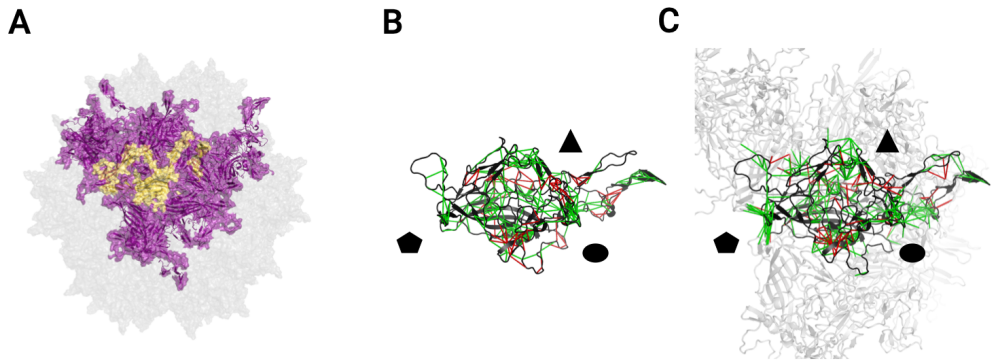



**Figure 4.2:** Analysis of AAV capsid using mutational frustration. AAV was analyzed using the Δ-Frustration Index metric, computed from applying the frustratometer tool to calculate mutational frustration of the virus capsid's monomer and multimer substructures. A: The AAV2 monomer (yellow) is defined as one VP3 subunit. The AAV2 multimer is defined as the central monomer (yellow) and the eight surrounding monomers (purple) containing a residue with any atom within a radius of 4.5Å around any of the central monomer residues. The rest of the AAV capsid (grey) was not included in this analysis. Mutational frustration of all possible residue interactions was computed using the frustratometer online tool for the AAV2 monomer and multimer. Minimally and highly frustrated interactions were plotted on Cryo-EM reconstructed structures of AAV2. B: Capsid residue interactions that are minimally (green) or highly (red) frustrated in the AAV2 monomer are indicated by connecting lines with the protein backbone in dark grey. C: Capsid residue interactions that are minimally (green) or highly (red) frustrated in the AAV2 monomer are indicated by connecting lines with the protein backbone in dark grey and other subunits in light grey. The capsid five-fold (pentagon), three-fold (triangle) and two-fold (circle) axes of symmetry are indicated.

have more neutral residues than a set of 314 monomeric proteins previously curated and analyzed from the Protein Data Bank (PDB), which have on average approximately 40% minimally frustrated residues and $>$ 10% highly frustrated residues [299, 300].

Single-residue mutational frustration indicies of the monomer and multimer assembly were combined to generate a Δ-frustration index. This change in frustration upon binding was previously computed for a benchmark of assemblies of homodimers and their subunits curated from the PDB [301]. 25% of residues generally found in protein-protein interaction domains exhibit an increase of approximately 1.5 single-residue mutational frustration index units (or, a Δ-frustration index of 1.5) in the bound state while 7% of residues exhibit a decrease in the single-residue mutational frustration index (or, a negative Δ-frustration index) [299]. However, in AAV2, the number of residues that undergo shifts in single-residue mutational frustration is much smaller than observed in the homodimeric proteins studied previously [299]. Approximately 6.0% of residues exhibit an increase in single-residue mutational frustration index in the multimer state greater than one unit (Δ-frustration index $>$ 1), while 3.7% of residues exhibit a decrease in single-residue mutational frustration index by more than one unit (Δ-frustration index $<$ 1)(Figure 4.3).

We hypothesized that this Δ-frustration index metric may highlight AAV capsid residues that shift in frustration upon subunit-subunit interaction, indicating that they play key roles in stability of either the monomer or the multimer assembly. By this logic, positive values of the Δ-frustration index indicate residues that favor the multimer assembly state, while negative values indicate residues that favor the monomer state. Interestingly, residues favoring the monomer state are prominent on the exterior surface of the capsid, while residues favoring the multimer state are proximate to the interior surface (Figure 4.4A). In the cross-section

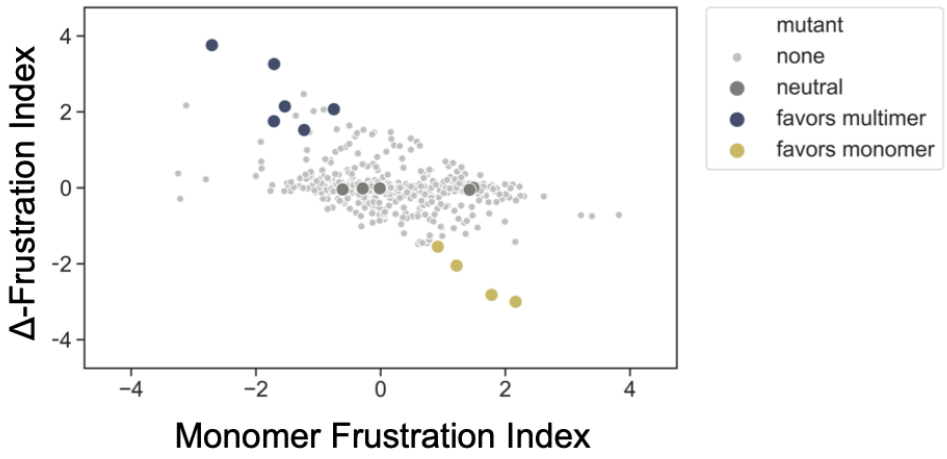



**Figure 4.3:** AAV2 Δ-Frustration Index plotted against the monomer frustration index. Residues selected for mutation are colored.

view, there are a high population of residues favoring the multimer as well as some favoring the monomer clustered around the capsid five-fold pore.

We also compared the Δ-frustration index to other residue-level parameters. There was no correlation with sequence conservation in the parvovirus family (Figure 4.4C). While there was no significant linear correlation with residue distance from the capsid center, residues favoring the monomer are positioned towards the exterior surface-exposed capsid domains, while residues favoring the multimer are positioned towards the interior capsid surface (Figure 4.4D). There is a slight positive correlation ($\rho$=0.13, p=0.002) between the number of inter-residue contacts, defined by a distance between the closest pair of atoms from each residue $< 4.5$ Å and the Δ-frustration index (Figure 4.4E). When these contacts are broken up into extra-subunit contacts (between pairs of residues found in different protein subunits) and intra-subunit contacts (between pairs of residues in the same subunit), there is a slight positive correlation between extra-subunit contacts and Δ-frustration index ($\rho$=0.12, p=0.006), but no significant correlation between intra-subunit contacts and Δ-frustration index (Figure 4.6). Single residue frustration index values were also computed for AAV serotypes 3-6, 8 and 9 using their solved structures. The Δ-frustration index is positively correlated between all serotypes studied (Figure 4.5)) with the degree of correlation corresponding to the evolutionary distance between serotypes [263].

A set of representative residues was selected based on the Δ-frustration index favoring the multimer, favoring the monomer, or neutral towards either state. Based on the threshold identified in [299] Δ-frustration index values $> 1.5$ were labeled as favoring the multimer. Δ-frustration index $< 1.5$ were labeled as favoring the monomer. Six residues from the 14 residues favoring the multimer were selected randomly, and all four residues favoring the monomer were selected. Five neutral residues were selected randomly from the 229 residues with Δ-frustration indices $> -0.1$ and $< 0.1$. Selected residues are depicted in Figure 4.4B. Selected residues were experimentally mutated to alanine to evaluate their role in capsid formation, thermal stability, and transduction efficiency.

In addition to frustration-based predictions, we used the associative memory, water mediated, structure and energy model (AWSEM), a coarse-grained protein force field with both physical and bioinformatic terms, to predict the change in energy (ΔE) of alanine mutants of selected residues in the monomer and multimer states. We developed a ΔΔE metric by subtracting the monomer ΔE from the multimer ΔE for each alanine

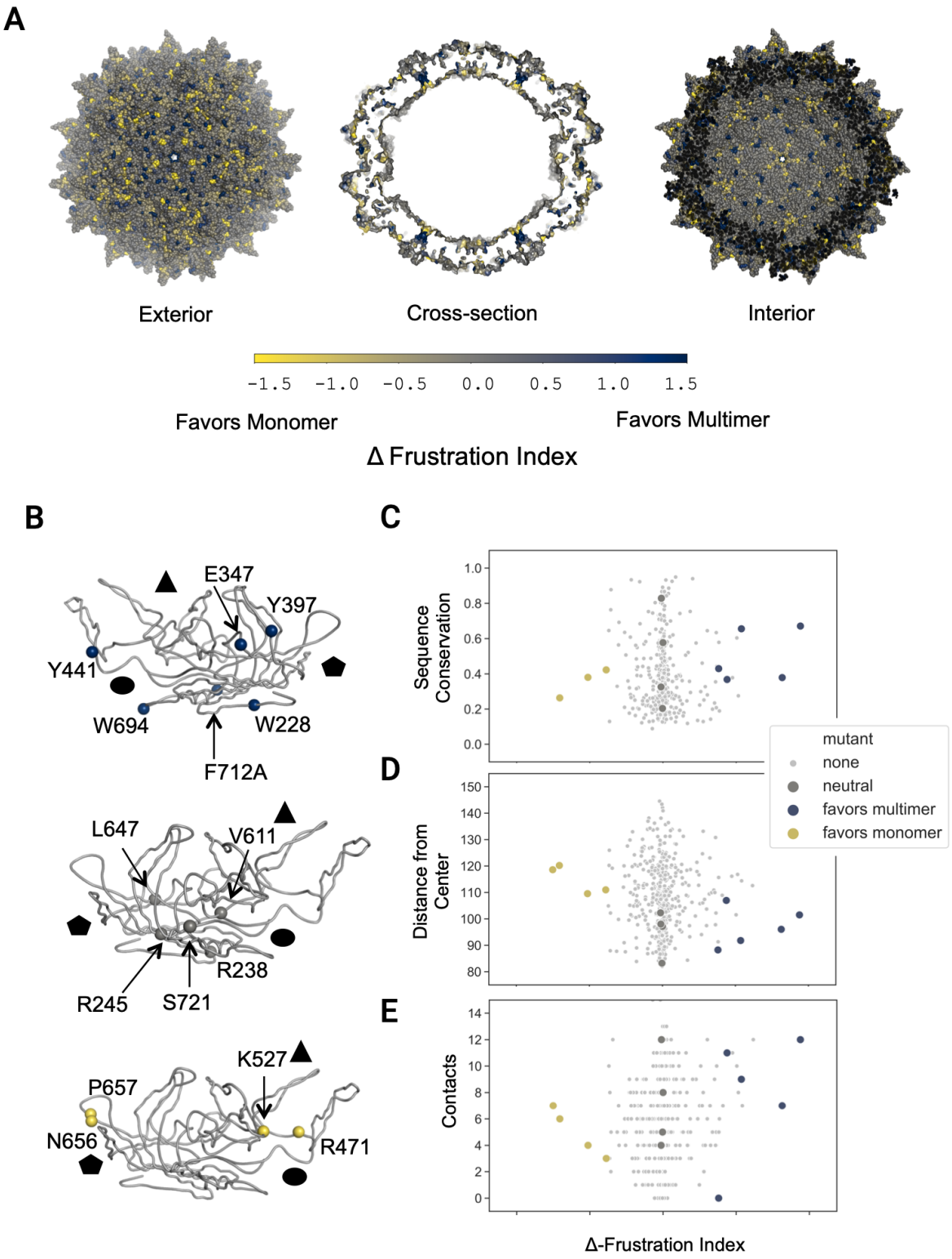



**Figure 4.4:** AAV2 Δ-Frustration Index and residues selected for experimental analysis. Caption continued on following page.

**Figure 4.4** ***(previous page)*:** AAV2 Δ-frustration Index was computed as the difference between the monomer and multimer frustration indices at each residue. Residues with a positive Δ-frustration Index are predicted to favor the multimer structure, while residues with a negative Δ-frustration Index are predicted to favor the monomer state. A: AAV2 Cryo-EM reconstructed 60-mer capsid structure colored according to Δ-frustration Index. View of the capsid from outside (exterior), center slice (cross-section), and inside (interior) are shown. B: Residues selected for experimental analysis. Residues favoring the multimer state (6 residues, blue, Δ-Frustration Index>1.5), residues favoring the monomer state (4 residues, yellow, Δ-Frustration Index<-1.5), and neutral residues (5 residues, gray, Δ-Frustration Index>-0.1 and <0.1) were selected for further study. The capsid five-fold (pentagon), three-fold (triangle), and two-fold (circle) axes of symmetry are indicated. Top subunit orientation shows the interior face of the VP3 subunit, whereas the middle and bottom subunit orientation show the exterior face. C-E: Δ-Frustration Index for AAV2 capsid residues are plotted against sequence conservation in the parvovirus family computed using shannon entropy of reweighted sequences and normalized from 0 to 1, where 1 indicates full conservation (C), distance of the residue from the center of the AAV2 capsid (D), and the number of contacts within 4.5Å of the residue (E). Residues selected for mutation are labeled.

mutant. As a more negative ΔE value indicates greater energetic favorability, alanine mutants with a lower ΔΔE score than wt are predicted to favor the multimer state, while alanine mutants with a higher ΔΔE score than wt are predicted to favor the monomer state. While frustration-based metrics make predictions about the properties of the native residue, the ΔΔE metric makes predictions about the alanine mutants that we generated experimentally. We examined the correlation of this alanine-specific stability metric and the Δ-frustration index with experimentally-derived measures of capsid formation and function.

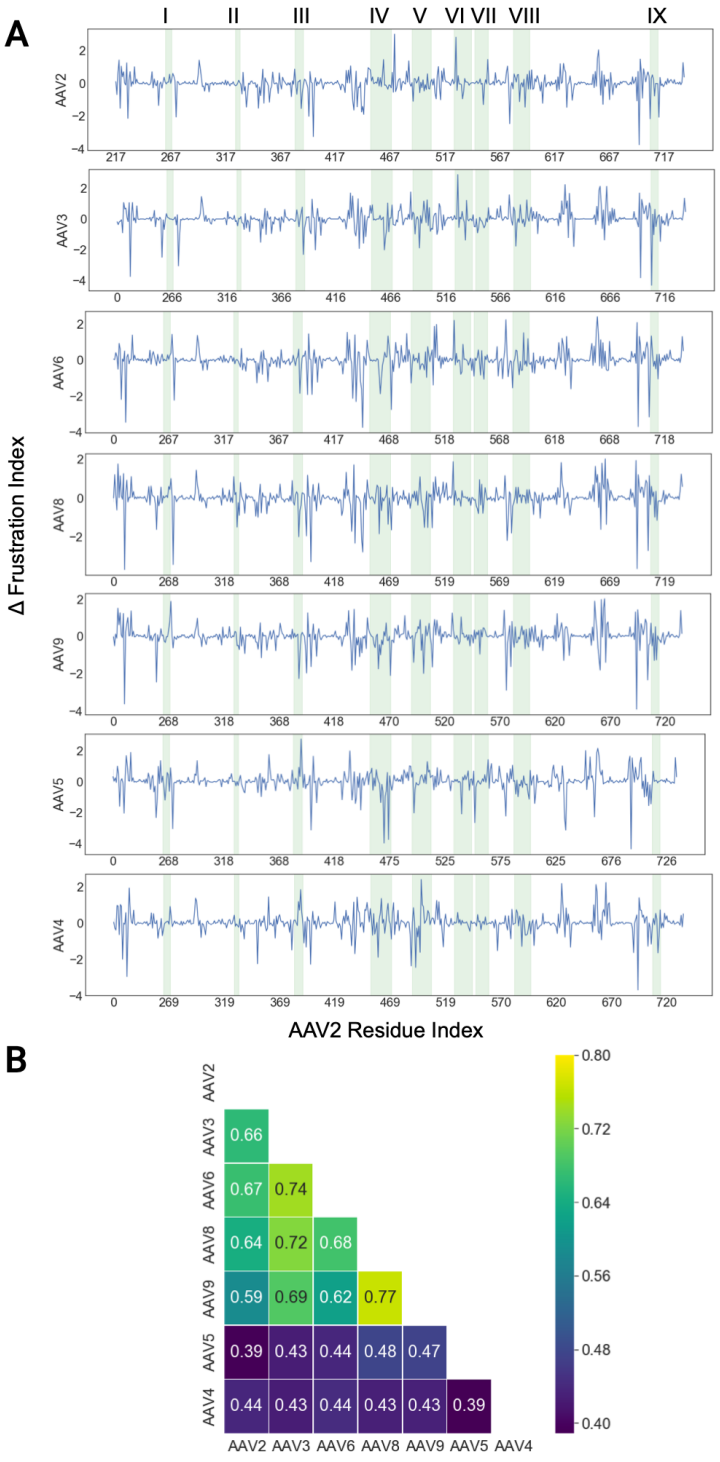



**Figure 4.5:** Δ-frustration indices for AAV serotypes 2, 3-6, 8-9. A) Δ-frustration indices are plotted against AAV2 residue numbering based on a clustal omega sequence alignment of these serotypes. Serotypes are organized according to phylogenetic relationships. B) Pearson correlations of the Δ-frustration indices between AAV serotypes.

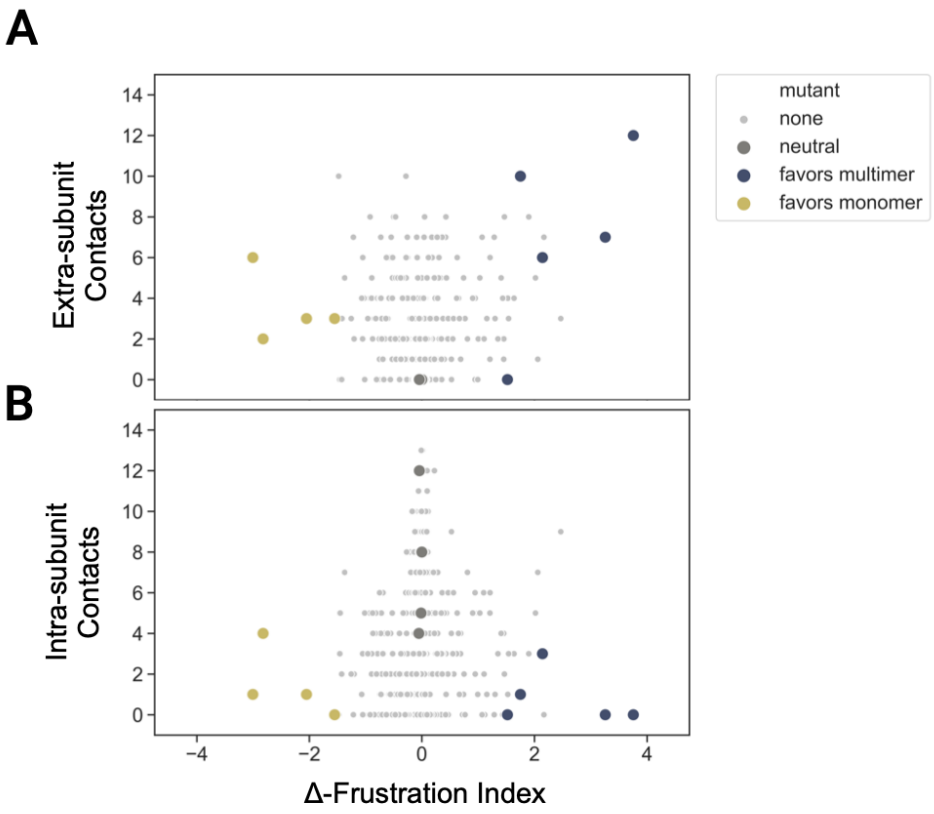



**Figure 4.6:** Plots of the AAV2 Δ-Frustration Index against the number of extra-subunit (A) and intra-subunit (B) contacts within 4.5Å of the residue.

### 4.2.2 Direct coupling analysis

We next applied direct coupling analysis (DCA) to make predictions about the fitness of the selected AAV2 alanine mutants. DCA is a global statistical model for capturing protein residue co-evolution derived from a multiple-sequence alignment of a protein family — in this case we used the capsid protein of parvoviruses drawn from Pfam (ID: PF00740) [287]. To evaluate this model's accuracy in identifying important AAV2 capsid residue-residue contacts based on parvovirus family sequences, we plotted predicted residue-residue contacts from DCA against known contacts defined as inter-residue distance $< 4.5$ Å in the AAV2 crystal structure (Figure 4.7A). The overlap between some DCA-predicted interactions and some contacts identified in the AAV2 capsid crystal structure, including contacts between residues that are distant in sequence space, indicates that DCA captures some structural contacts as interacting pairs, so this analysis may also be able to make accurate predictions about interacting residue pairs that are not structural contacts in the AAV capsid.

We also investigated the DCA-derived Hamiltonian scores of the AAV2 capsid alanine mutants ($H_{DCA(Ala)}$). DCA generates a joint probability distribution that describes the probability that a given sequence is a member of the characterized sequence family. This probability is log transformed to produce a unitless quasi-energy hamiltonian [289, 302]. Lower values of the hamiltonian correspond to sequences that are more likely to belong to the characterized sequence family and thus more likely to be structurally and functionally

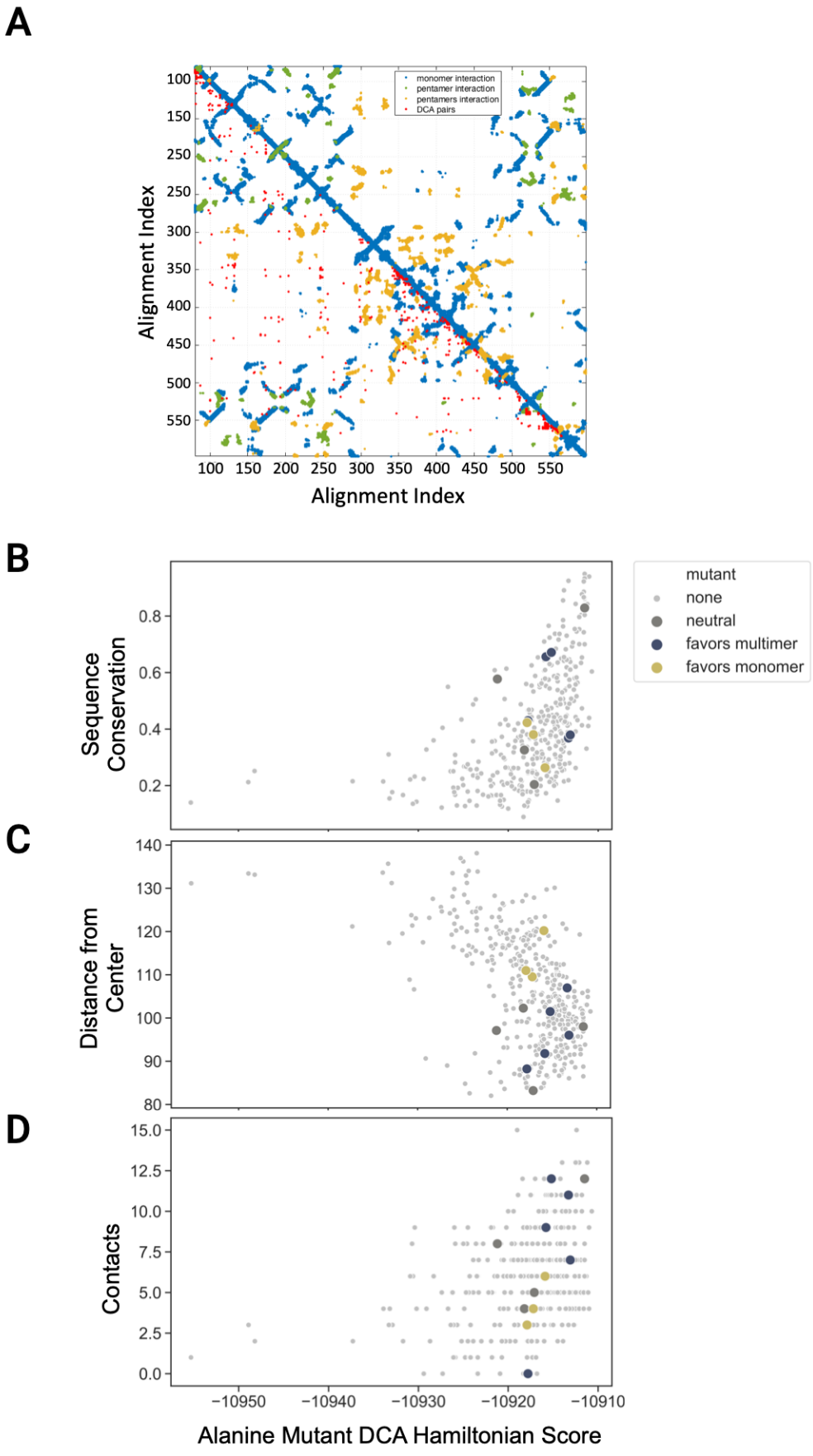



**Figure 4.7:** AAV2 capsid residue-residue contact prediction by direct coupling analysis (DCA) and comparison of $H_{DCA(Ala)}$ to other residue parameters. Caption continued on following page.

**Figure 4.7** ***(previous page)*:** DCA was applied to a multiple sequence alignment of sequences identified as likely parvovirus coat proteins in NCBI. A: Predicted residue-residue contacts from DCA and residue-residue contacts identified in the AAV2 capsid crystal structure are plotted on a contact map. DCA-predicted contacts (red), structurally-determined monomeric contacts (blue), and structurally-determined multimeric contacts (green and yellow) are plotted at the intersection of the two contacting residue's AAV2 capsid sequence indices. B-D: Plots of the $H_{DCA(Ala)}$ score against sequence conservation measured in the parvovirus family (B), distance of the residue from the center of the AAV2 capsid (C), and the number of contacts within 4.5Å of the residue (D). Residues selected for mutagenesis are highlighted (large circles). Other residues are depicted as small grey circles.

similar to the members of the sequence family. This hamiltonian term was computed for AAV2 sequences with each residue position individually mutated to alanine and compared to other residue-level parameters.

$H_{DCA(Ala)}$ is correlated with sequence conservation within the parvovirus family($\rho$=0.52, p<<0.001). DCA generates a global model of the parvovirus sequence family and makes predictions of mutant fitness based on the prevalence of residue identities and residue pairs identified from the sequence family. Therefore, DCA is likely to predict that mutations of a residue that is highly conserved will be detrimental to fitness, so this correlation is expected. While there is no linear correlation between $H_{DCA(Ala)}$ and residue distance from the capsid center, residues at extremes of distance (exposed towards the interior face of the capsid or the exterior surface) have a lower $H_{DCA(Ala)}$, suggesting that alanine mutations of these residues are more likely to be tolerated than alanine mutations of residues buried in the capsid shell. $H_{DCA(Ala)}$ is significantly correlated with residue-residue contacts ($\rho$=0.35, p<<0.001), suggesting that DCA is able to accurately identify residue interactions in capsid proteins. When these contacts are broken up into extra- and intra-subunit contacts, there is a positive correlation between intra-subunit contacts and $H_{DCA(Ala)}$ ($\rho$=0.31, p<<0.001), but no significant correlation between extra-subunit contacts and $H_{DCA(Ala)}$ (Figure 4.8).

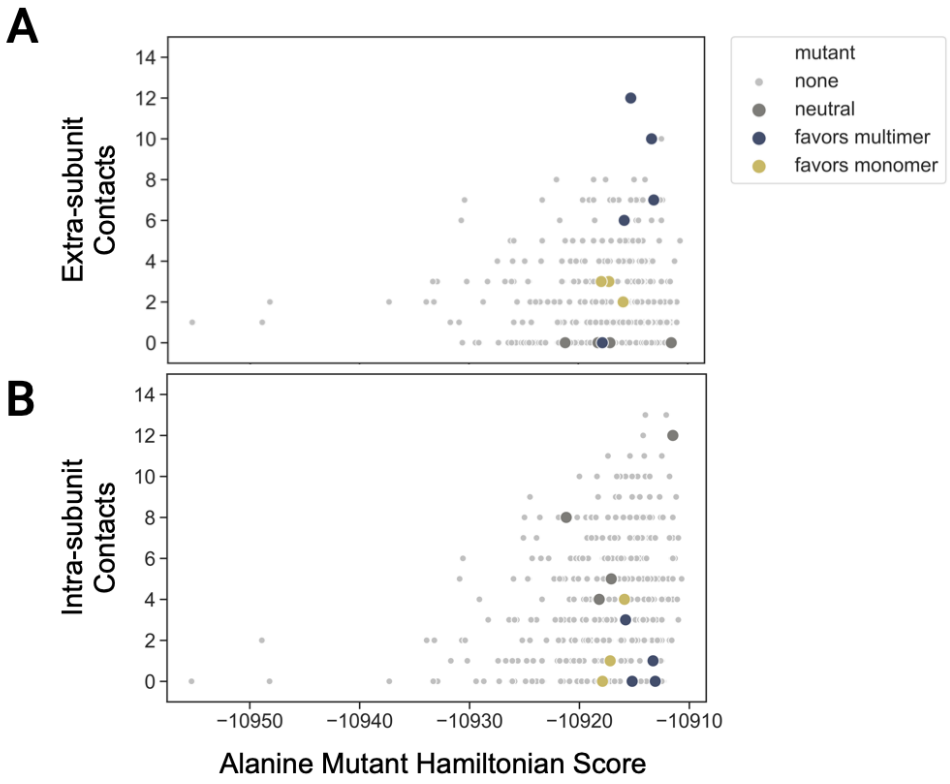



**Figure 4.8:** AAV2 $H_{DCA(Ala)}$ plotted against the number of extra-subunit (A) and intra-subunit (B) residue-residue contacts within 4.5Å of the residue.

### 4.2.3 Formation of AAV2 capsid alanine mutants

We mutated the selected AAV2 capsid residues to alanine and generated genome-packaging viruses (Figure 4.9A). Among mutants of residues favoring the multimer, four of six mutants exhibited decreased genomic titers, while the remaining mutants exhibit titers comparable to wild-type (WT) AAV2 capsid. Among mutants of residues favoring the monomer, one mutant (P657A) exhibits decreased genomic titers, while the other three exhibit titers comparable to WT. Among mutants of neutral residues, two of five mutants exhibited decreased genomic titer, while three exhibit titers comparable to WT.

We first plotted the virus production titers against basic metrics derived from the capsid structure and an alignment of parvovirus family sequences. Mutant capsid assembly is significantly negatively correlated with the number of residue-residue contacts made by the native residue in the capsid structure, *i.e.* the greater the number of native contacts a residue has, the lower the yield of virus formed when that residue is mutated to an alanine (Figure 4.9B). Additionally, virus production is significantly positively correlated with the frequency of alanine residues observed at the native residue's position in a multiple sequence alignment of parvovirus family sequences (Figure 4.9C). In other words, if an alanine is found in a homologous capsid position in many other parvoviruses, the alanine mutation in the AAV2 capsid is better tolerated and yields higher virus production levels.

Mutant virus formation is not significantly correlated with the native residue's Δ-Frustration Index (Figure 4.9D), the predicted ΔΔE of alanine mutants computed with AWSEM-MD (Figure 4.9E), or DCA-based $H_{DCA(Ala)}$ scores (Figure 4.9F). However, there appears to be a general trend where mutants of

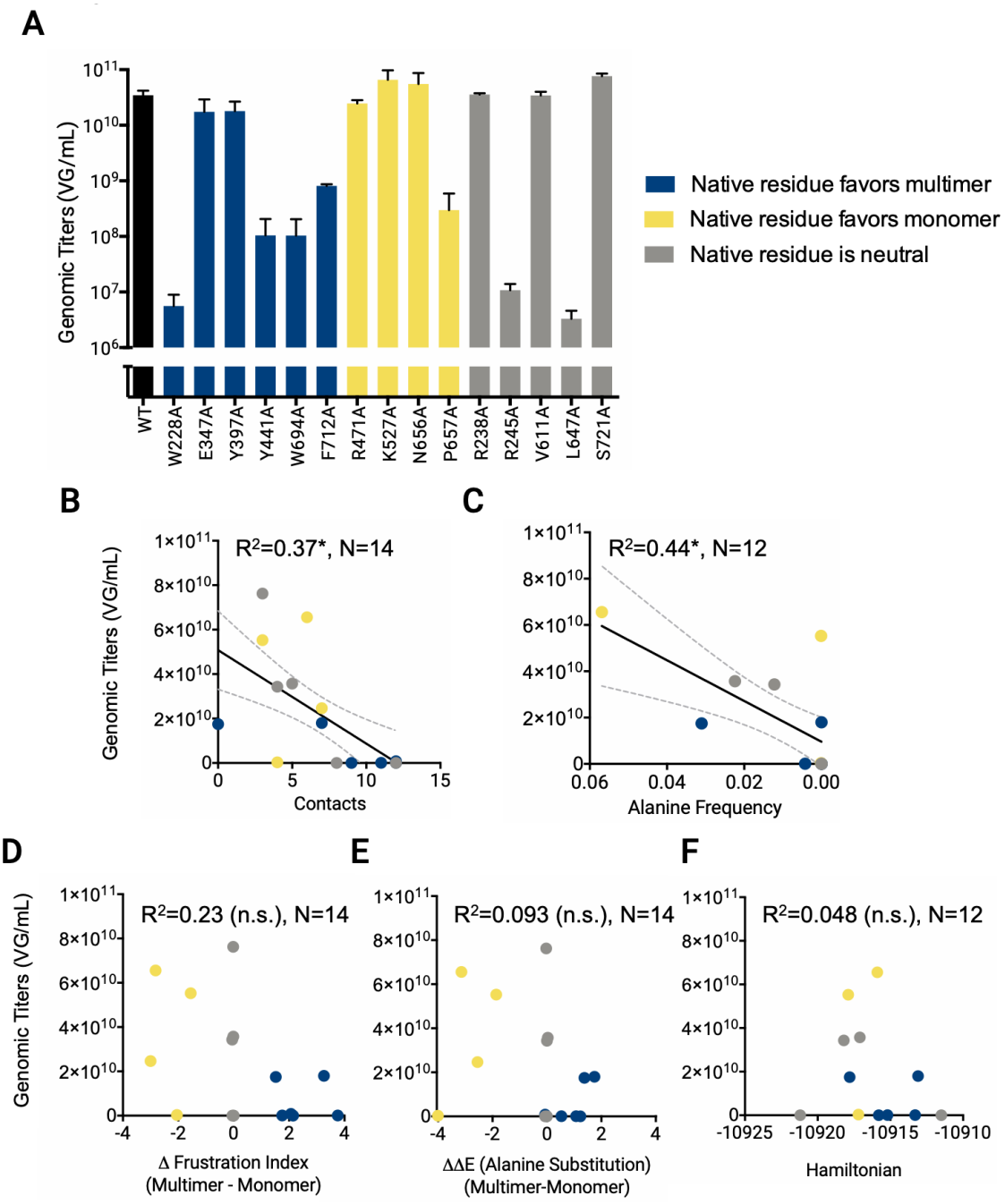



**Figure 4.9:** Capsid assembly of alanine mutants. Selected residues were mutated to alanine and viruses produced. Genomic titers of one-plate virus preps were determined through qPCR. A: Genomic titers of alanine mutants expressed as viral genomes per mL (VG/mL). Mutants of residues favoring the multimer are in blue, mutants of residues favoring the monomer are in yellow, and neutral residues are in gray. N=3 independent virus preps. Error bars are SEM. One-way ANOVA was performed with Dunnett's post-hoc multiple comparison test to compare mutants to WT - no significant differences were detected. B-F: Genomic titers are plotted against number of residue-residue contacts in the native capsid structure (B), alanine frequency in the parvovirus sequence family (C), Δ-Frustration Index (D), ΔΔE of alanine substitutions (E), and $H_{DCA(Ala)}$ (F). B-F: N=14 for structure-based analyses, N=12 for sequence based analyses as some residues were excluded due to a high-gap frequency at their alignment position. Regression lines with 95% confidence interval are plotted for significant correlations. $R^2$ values and significance levels are reported (*p<0.05, **p<0.01, ***p<0.001).

residues with higher Δ-frustration Indices have higher ΔΔE energies and reduced formation. Proline has a unique secondary structure that can often result in challenges with predicting backbone angles in molecular dynamics models [303, 304]. If the P657A mutant is excluded from this analysis, then virus formation is significantly correlated with Δ-Frustration Index of native residues ($R^2$=0.36*, N=13) and ΔΔE of alanine mutants ($R^2$=0.30*, N=13) but not $H_{DCA(Ala)}$. The results thus far demonstrate that both the number of residue-residue contacts and metrics derived from structure-based models (Δ-frustration index and AWSEM-derived ΔΔE) are somewhat related to capsid formation, but neither metric fully predicts assembly.

### 4.2.4 Thermal stability of AAV2 alanine mutants

Viruses that formed with sufficient titers for further analysis (>$10^{10}$ VG/mL) were screened for thermal stability through a genomic protection assay. Specifically, viruses were incubated at a range of temperatures near the WT AAV2 melting temperature, which was previously reported as 72.4 °C [97]. The virus samples were then treated with a nuclease to degrade any uncoated viral genomes. After nuclease inactivation, the samples were then assayed for number of remaining genomes (*i.e.* genomes that are protected by an intact capsid). Capsids that are thermostable at a particular temperature would effectively protect their genomes against nuclease digestion. Most capsids exhibit thermal stability generally comparable to WT, with the exception of the R471A and K527A mutants. These two mutants appear to lose genomic protection after incubation at 66 °C, the lowest incubation temperature tested (Figure 4.10A). To further characterize these mutants' reduction in thermal stability, the melting points of these mutants were determined using a differential scanning fluorescence assay (Figure 4.11A). R471A exhibits a 6.8 °C decrease in melting temperature and K527A exhibits a 3.5 °C decrease in melting temperature (Figure 4.11B).

Thermal stability at 68 °C was compared to various computational metrics of fitness as described above. Virus thermal stability is not correlated with the number of residue-residue contacts made by the mutated residue (Figure 4.10B) or with the frequency of alanine in an alignment of parvovirus sequences (Figure 4.10C). Thermal stability is, however, correlated with both the Δ-frustration index and the ΔΔE of alanine mutants predicted with AWSEM-MD. Interestingly, mutants of residues that are predicted to favor the monomer by Δ-frustration index have lower genomic protection at 68 °C (Figure 4.10D). Similarly, alanine mutants predicted to have a lower ΔΔE between the multimer and monomer than WT, and thus predicted to favor the multimer state more than WT, have lower genomic protection at 68 °C (Figure 4.10E). $H_{DCA(Ala)}$ shows no correlation with capsid thermal stability (Figure 4.10F). These results indicate that metrics derived from structure-based models (Δ-frustration index and AWSEM-derived ΔΔE) may be useful in predicting deficiencies in thermal stability by identifying capsid mutants that favor the multimer structure over the monomer as compared to WT.

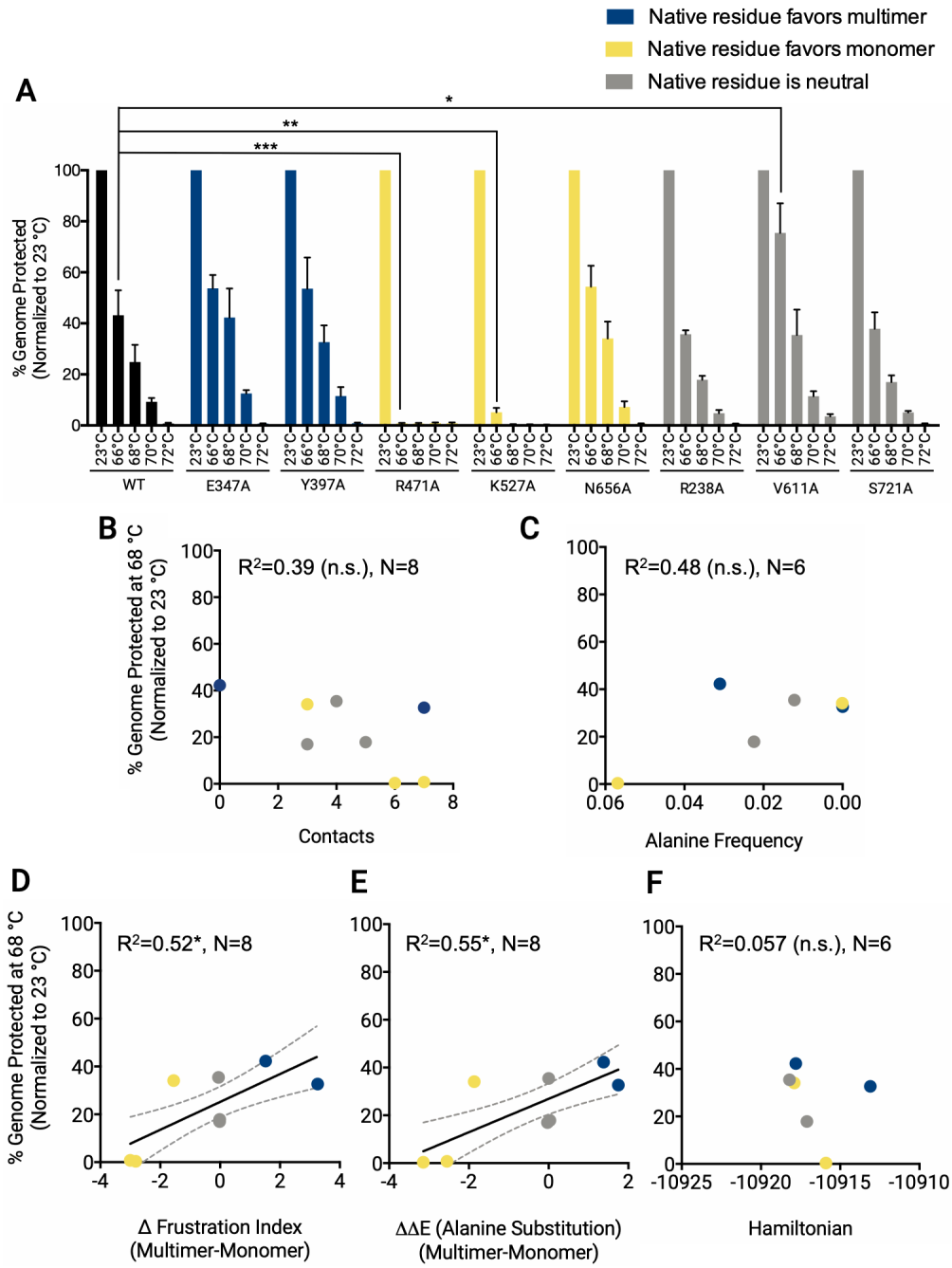
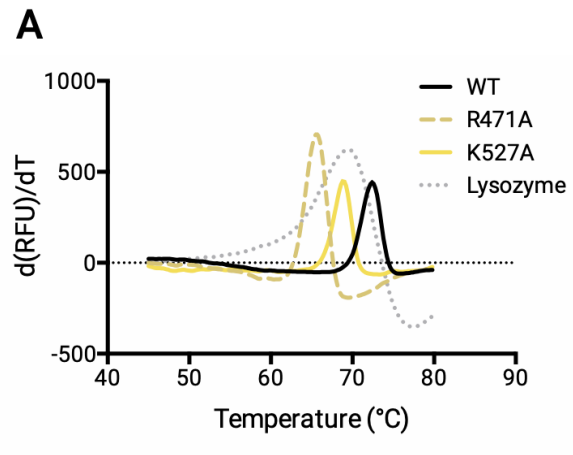



**Figure 4.10:** Thermal stability of alanine mutant capsids. Viruses were exposed to various temperatures and then assayed for genomic protection. Temperature-treated samples were incubated with benzonase or sham buffer and genomes quantified using qPCR. A: Genomic protection after incubation at elevated temperatures. N=3 independent experiments titered in duplicate. Error bars are SEM. Two-way ANOVA was performed with Dunnett's post-hoc multiple comparison test to compare mutants to WT at each tested temperature. Percent genome protected is the fraction of genomes in the benzonase-treated sample as compared to the sham-treated sample. B-F: Genomic protection at 68 °C plotted against number of residue-residue contacts in the native capsid structure (B), alanine frequency in the parvovirus sequence family (C), Δ-Frustration Index (D), ΔΔE of alanine substitutions (E), and $H_{DCA(Ala)}$ (F). Regression lines with 95% confidence interval are plotted for significant correlations. $R^2$ values and significance levels are reported. *p<0.05, **p<0.01, ***p<0.001.




B

| Capsid | Melting Temperature (°C) |
|---|---|
| WT AAV2 | 72.4±0.2 |
| R471A | 65.6±0.2 *** |
| K527A | 68.9±0.1 *** |

**Figure 4.11:** Melting temperature of selected alanine mutant capsids R471A and K527A. The melting temperature of selected capsids with reductions in thermal stability were determined using differential scanning fluorescence. A: Derivative in fluorescence intensity with heating of capsids. Fluorescence increases upon the binding of Sypro Orange to exposed hydrophobic pockets as capsids denature. Traces are the average of N=3 independent experiments. Lysozyme is included as a control. B: Melting temperature of capsids determined as the peak of plots shown in panel A. N=3 independent experiments in duplicate, SEM reported. One-way ANOVA was performed with Dunnett's post-hoc multiple comparison test to compare melting temperatures of mutants to WT. ***$p<0.001$

### 4.2.5 Transduction efficiencies of AAV2 capsid alanine mutants

Lastly, viruses were also screened for their ability to transduce HEK293T cells. Mutants of residues favoring the multimer and residues favoring the monomer show a wide range of transduction levels, with some mutants severely deficient in transduction and others comparable to WT (Figure 4.12A). Mutants of neutral residues all showed transduction levels similar to WT. The number of residue-residue contacts (Figure 4.12B), the Δ-Frustration Index (Figure 4.12D) and the ΔΔE of alanine substitutions (Figure 4.12E) are not correlated with virus transduction ability. As these models are derived from viral capsid structure, they do not incorporate other interactions the AAV capsid must successfully undergo with its cellular environment to transduce cells. The frequency of alanine in an alignment of parvovirus sequences is also not correlated with transduction (Figure 4.12C). The $H_{DCA(Ala)}$ score, however, is highly correlated with virus transduction (Figure 4.12F). Capsid alanine mutants that are predicted to have higher fitness in the DCA-derived model of parvovirus coat protein sequences are more successful at genome delivery into host cells.

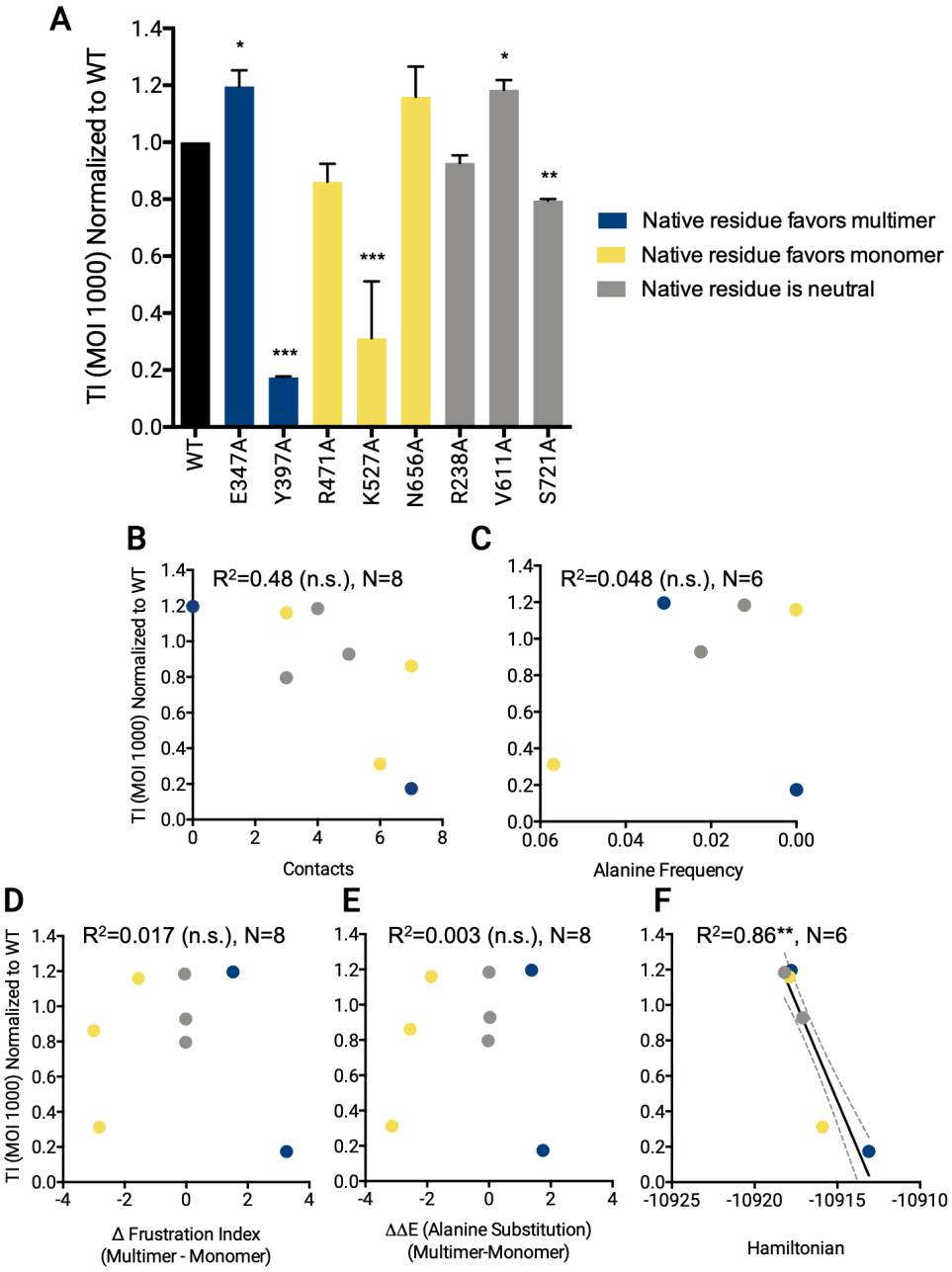



**Figure 4.12:** Transduction ability of alanine mutant capsids. HEK293T cells were transduced with viruses packaging scGFP transgene at an MOI of 1000, and transgene expression was quantified at 48h post-transduction using flow cytometry. A: Transduction index of viruses. Transduction Index (TI) is the product of % GFP+ cells and the geometric mean fluorescence intensity, the metrics obtained from flow cytometry. N=3 independent experiments done in triplicate. Error bars are SEM. One-way ANOVA was performed with Dunnett's post-hoc multiple comparison test to compare mutants to WT. B-F: TI of alanine mutants plotted against number of residue-residue contacts in the native capsid structure (B), alanine frequency in the parvovirus sequence family (C), Δ-Frustration Index (D), ΔΔE of alanine substitutions (E), and $H_{DCA(Ala)}$ (F). Regression lines with 95% confidence interval are plotted for significant correlations. $R^2$ values and significance levels are reported. *p<0.05, **p<0.01, ***p<0.001.

## 4.3 Discussion

We applied two computational approaches to infer energy functions for AAV based on either a coarse-grained structural prediction model or a global statistical model of sequences in the parvovirus family. We generated energies from these models and tested a series of AAV2 mutants to characterize the relationship between these predicted energies and viral formation and function.

Single-residue frustration analysis of the AAV2 monomer structure indicated more neutral residues than previously observed in the monomer domains of dimerizing proteins. Additionally, fewer residues undergo shifts between the monomer and multimer state than previously observed in dimerizing proteins, and the shifts are not particularly skewed towards minimal frustration in the assembly. Molecular dynamics models of virus assembly have previously shown that interactions between viral subunits must be relatively weak to permit viral assembly while avoiding kinetic traps due to malformed structures — this property may result in a less dramatic shift towards minimal frustration upon binding than observed in smaller protein assemblies [305, 306]. Residues that are more minimally frustrated in the multimer tend to occur on the interior face of the capsid and have more contacts with other residues, suggesting that essential binding interactions take place in these domains that are conserved across the parvovirus family [307]. Conversely, residues that are more minimally frustrated in the monomer are more distal and surface exposed in the capsid structure, and have less contacts with other residues. These residues may play a role in capsid interactions with other proteins on the virus's transduction pathway as opposed to intra-capsid interactions. Indeed, mutation of R471 has previously been shown to reduce immune system interactions and K527 is known to be proximal to the AAV2 heparin binding pocket, although mutation does not impact heparin binding [51]. The five-fold pore contains residues that show large shifts in $\Delta$-Frustration in both directions. The five-fold pore is known to facilitate virus genome packaging and extrusion of the capsid N-termini in response to the endosomal environment, so it likely is a key point for structural transformations in the capsid throughout its life cycle [43].

Direct coupling analysis produced interesting findings, despite the paucity of available parvoviral sequences. The direct coupling analysis joint probability distribution was inferred from an alignment of parvoviruses with 125 effective sequences ($M_{eff}$), 2.2% of the 5577 starting sequences after re-weighting sequences to account for sequence homology. A typical $M_{eff}$ is 1/2-1/3 of the starting count — however this lower yield is to be expected as only 181 parvovirus sequences have been identified and the high number of sequences in pfam are likely derived from isolates with small sequence variations [287,308]. While this $M_{eff}$ is lower than the required number previously postulated to produce high rates of true positive contacts, the inferred model shows promise in identifying both intra- and inter-subunit contacts in the AAV2 structure. Predictions of $H_{DCA(Ala)}$ are moderately correlated with each residue's number of contacts, again indicating that the model has captured direct relationships between residues. $H_{DCA(Ala)}$ has a significant correlation with intra-subunit contacts but not with extra-subunit contacts, which may be related to the strong conservation of the structural beta-barrel motifs within each subunit in the parvovirus family [180]. While the $\Delta$-Frustration metric appears to have a linear relationship with residue distance from the capsid center, the $H_{DCA(Ala)}$ metric is low for both the closest and farthest residues from the capsid center. This indicates that in the $H_{DCA(Ala)}$ model, residues buried within the protein shell are predicted to be less tolerant to alanine mutation than interior- or exterior-surface exposed residues.

Formation of alanine mutant virions has a small correlation with both the number of residue contacts observed in the capsid structure and the frequency of alanine residues in the parvovirus sequence alignment. These results suggest that basic structural contact information can provide information about complete virion formation, as well as residue frequency information drawn from a family of related sequences. The

Δ-frustration index and ΔΔE alanine substitution metric are significantly correlated with genomic titers only if the P657A mutant is excluded from analysis, perhaps due to proline's difficult-to-model secondary structure. The $H_{DCA(Ala)}$ metric shows no correlation with viral formation. From this data, it is unclear whether computational models provide more insight about viral formation than basic metrics derived from known information about capsid structure and the parvovirus sequence family. No model appears to capture formation fully, perhaps because these approaches do not consider any of the other known AAV and helper virus proteins required for capsid assembly and viral genome packaging [6, 32, 34, 41]. These models may potentially perform better in prediction of AAV2 VP monomer formation or capsid protein assembly as opposed to complete genome-containing virion formation, as these intermediate steps in assembly require fewer interactions with protein co-factors and genomic DNA [23].

While the majority of mutant capsids tested show no notable deficiencies in thermal stability, the R471A and K527A mutants show significant decreases in melting temperature. This result is somewhat surprising, as the Δ-Frustration metric indicates that these two residues are more minimally frustrated in the monomer, so mutations would be expected to destabilize the monomer more than the multimer. The ΔΔE alanine substitution metric also predicts that these two mutant residues would result in increased stability of the capsid assembly and decreased stability of the monomer. However, this result may may be due to the stability of the capsid monomer playing a larger role in overall capsid thermal stability. Indeed, AAV capsids are believed to melt at high temperatures due to denaturation of the capsid cooperatively, as capsid melting temperatures are much higher than monomer melting temperatures [267]. This theory may be evaluated by testing the melting point of mutant VPs synthesized in bacterial expression systems without other required components for capsid assembly. Overall, AWSEM-derived metrics appear to have potential for identifying residues that impact capsid melting temperature.

AAV transduction is a complex multi-stage process, requiring the capsid to interact with extracellular receptors, escape the endosome, traffic to and enter the cell nucleus, and disassemble to release the transgene for transcription [309]. The $H_{DCA(Ala)}$ metric is highly correlated with transduction. As the DCA approach is based on a global model drawing on sequences from the parvovirus family, this model may better be able to predict which mutated sequences are able to conduct needed environmental interactions to traverse the virus's infectivity pathway. Furthermore, pairwise residue-residue interactions appear to be essential to this result, as the parvovirus family alanine frequency metric is not correlated with transduction. However, this result is derived from the data of only six mutants. Eight mutants formed with titers sufficient for viral transduction experiments, but the parvovirus alignment used for this study does not contain sufficient non-gap sequences at positions R471 or S721 so $H_{DCA(Ala)}$ was not computed for mutants of these residues. The DCA approach could be applied to a larger capsid single-mutant library using next-generation sequencing technology to further validate this preliminary finding [262]. Additionally, DCA may be tested for its ability to capture the performance of multiple-residue mutants.

AWSEM and DCA energy predictions were previously found to be highly correlated for protein monomers, so the divergence in results observed in this study was surprising [289]. In the mutants analyzed, the AWSEM monomer ΔE prediction actually has a high negative correlation with DCA, while there is no significant correlation between the AWSEM multimer ΔE or the $\Delta\Delta E$ metric (Table 4.1). This difference may be attributable to the many roles of viral proteins in transduction. This may also be a function of viral capsid metastability, as viruses must maintain stable structures to protect their genomes but then fall apart once the cargo arrives at its intended intracellular destination [305].

**Table 4.1:** Pearson correlations between $H_{DCA(Ala)}$ and monomer $\Delta E_{Ala}$, multimer $\Delta E_{Ala}$, and $\Delta\Delta E_{Ala}$ computed with AWSEM-MD. N=12 p<0.05:*, p<0.01:**, p<0.001:***

| | **monomer ΔE** | **multimer ΔE** | $\Delta\Delta E$ |
|---|---|---|---|
| **DCA Hamiltonian** | -0.81** | 0.24 (*n.s.*) | 0.053 (*n.s.*) |

## 4.4 Conclusion

We have explored computational approaches for translating information about the AAV2 capsid protein into predictions about viral mutant function. DCA has strong potential for predicting viral transduction, while the frustratometer and predictions of the free energy of alanine mutants using AWSEM show some promise in predicting viral stability. Both AWSEM and DCA capture features of the protein energy landscape — while AWSEM directly models thermodynamics of the protein, DCA incorporates information from the evolutionary pressures that have shaped all members of the sequence family to fold robustly and form functional proteins. However, DCA also incorporates other constraints on evolution for protein function as well as formation and stability. These two approaches have demonstrated similar trends in prior studies, but in this case they appear to capture different information about capsid stability and function. This may be due to the complex energy landscape of the metastable viral capsid or to the large number of protein-protein interactions AAV encounters in its infectivity pathway.

In future studies, the strengths of these two approaches may be combined to create a more complete model of AAV. AWSEM may be better able to capture AAV capsid metastability by considering the virus's monomer and assembly states as dual energetic basins in a structure-based model [310]. DCA-derived contacts may be used to predict structure and capsid interactions for the unresolved AAV VP1 and VP2 N-terminal domains, which package inside the AAV capsid and then exhibit dynamic extrusion in response to the endosomal environment during transduction [9]. Combined approaches using DCA to feed likely contacts into structure-based thermodynamic models have identified binding domains of heterodimer pairs [311, 312] — this strategy may prove useful in identifying pairwise interactions between the AAV capsid and other known binding partners, including those derived from the AAV genome, from helper viruses, and from host cells. To validate trends noted here and refine the application of these models to AAV, larger-scale mutational studies exploring a variety of mutation types are needed. This approach will ideally lead to the development of an accurate *in silico* model of AAV, incorporating all interactions within the capsid and with other proteins required for successful gene delivery. This model can then be used to pre-screen engineered viruses, accelerating the development of optimized vectors for safer and more effective gene therapy.

## 4.5 Experimental

### 4.5.1 Identifying residue contacts within and between AAV2 capsid subunits

The high-resolution crystal structure of AAV2 (PDB ID: 1lp3) was visualized and analyzed in PyMol 2.2.0. Residue contacts were identified as any pair of residues containing atoms within 4.5 Å of each other. Intra-residue contacts were identified within an AAV2 subunit, while inter-residue contacts were identified between AAV2 subunits.

### 4.5.2 Frustratometer analysis of AAV capsid subunits

Using the crystal structure of AAV2, two types of constructs were created for analysis: a monomer form, consisting of a single AAV2 capsid subunit, and an assembly form, consisting of a central monomer and each monomer containing a residue with any atom within a radius of 4.5 Å around any of the central monomer residues. The AWSEM-MD Frustratometer Server* with default parameters (sequence separation = 12 and no electrostatics) was used to calculate the single-residue frustration index (F) of the monomer and assembly structures [272]. The ΔF index was then calculated for each residue as $F_{assembly}$-$F_{monomer}$ [299]. This analysis was repeated for AAV serotypes 3b, 4, 5, 6, 8, and 9 (Table 4.2).

| Serotype | PDB ID |
|---|---|
| AAV1 | From Agbandje-McKenna group |
| AAV2 | 1lp3 [307] |
| AAV3b | From Agbandje-McKenna group |
| AAV4 | 2g8g [313] |
| AAV5 | 3ntt [314] |
| AAV6 | 3oah [315] |
| AAV7 | 2qa0 [316] |
| AAV7 | 3ux1 [317] |

**Table 4.2:** Structures of AAV serotypes used for computational analyses [300].

### 4.5.3 Computational alanine mutagenesis using AWSEM

To calculate the predicted change in energy between the monomer and multimer states for alanine mutants, we applied the Associative memory, Water mediated, Structure and Energy Model (AWSEM), a predictive protein coarse-grained model that combines terms from energy landscape theory and information from a database of known protein structures. This integration of force-field modeling with bioinformatic protein data allows for a powerful tool that can predict changes in protein structure. AWSEM is a learning algorithm that uses training proteins to develop its prediction of the force field governing protein folding. This model uses 3 beads, or spatial coordinates with physical properties, to represent each residue in its coarse graining. The physical portion of the model focues on hydrogen bonding, hydrophobic interactions and water-mediated interactions between hydrophilic residues. Bioinformatic data is used to bias the structure of sequence fragments (9 residues or less) towards the structure of similar sequence fragments found in other proteins. The energy function of this model is described in Equation 4.1.

$$V_{AWSEM} = V_{backbone} + V_{contact} + V_{burial} + V_{helical} + V_{FM} \tag{4.1}$$

The backbone energy ($V_{backbone}$) incorporates information about protein chain structure. The contact energy ($V_{contact}$) describes tertiary residue interactions between residues that are at least 10 amino acids

*http://www.frustratometer.tk/

apart, factoring in information about whether water or protein is the dominant factor in the local environment. The burial energy ($V_{burial}$) confers information about the tendency of residues to be buried (contacting protein) or surface-exposed (contacting environment). The helical energy term ($V_{helical}$) captures hydrogen bonding that takes place in helical structures. Finally, the fragment memory term ($V_{FM}$) factors in information from the RCSB PDB database about the structure of similar sequence fragments [318]. This fragment memory term was used to predict changes in the structure of the AAV monomer and multimer upon alanine substitution. The $\Delta\Delta E$ of alanine substitution was then computed as described in Equation 4.2.

$$\Delta\Delta E_{ala} = V_{AWSEM_{Multimer}} - V_{AWSEM_{Monomer}} \tag{4.2}$$

### 4.5.4 Alanine frequency in the parvovirus sequence family

NCBI sequences for the Parvovirus VP2 coat protein multiple sequence alignment developed using a hidden markov model were extracted from Pfam (PF00740). The alignment contained 5577 sequences. Sequences with gaps greater than 100 base pairs were dropped, and sequences with greater than 90% identity were re-weighted as described previously [287], leaving 125.91 effective sequences. Columns in the alignment with less than 40% non-gapped values were dropped from the analysis due to sparse data. The frequency of alanine at each position was then calculated.

### 4.5.5 Direct coupling analysis of AAV2 capsid

Direct coupling analysis was applied to the preprocessed alignment described above to infer a global statistical model of co-evolved residue interactions. DCA infers a joint probability distribution based on a maximum entropy approach, with terms for single-position values (fields) as well as pairwise couplings [287]. Pairwise couplings may be restricted to occur within a certain spatial residue distance based on the structure of a given member of the sequence family [289]. From this distribution, the direct information parameter may be computed for sequence pairs, providing the amount of mutual information attributable specifically to the direct coupling between each pair [287]. Identified pairs above a certain threshold were plotted against measured residue-residue contacts from the AAV2 structure to assess this approach's ability to identify residue interactions. The joint probability distribution was also used to calculate the probability $P_{DCA}$ that a given sequence is a member of the characterized sequence family, and this probability was converted into a unitless energy hamiltonian $H_{DCA} = \log P_{DCA}$ [289, 302]. This hamiltonian term was computed for AAV2 sequences with each residue position mutated to alanine.

### 4.5.6 Site-directed mutagenesis of AAV2 cap gene

Site-directed mutagenesis of the AAV2 cap gene was performed to substitute selected residues with alanines. The pXX2 plasmid containing the wtAAV2 rep and cap genes was used as the template [319]. Primers containing the desired alanine mutations were designed using the QuikChange Primer Design Program[†] and purchased from Integrated DNA Technologies, Inc. 18 cycles of PCR amplification were conducted according to the QuikChange protocol using Pfu Ultra polymerase (Agilent Technologies). After cycling was complete, template DNA was removed by digesting with DpnI (New England Biolabs). Resulting plasmids were sequence-verified through an external vendor (Genewiz).

[†]https://www.genomics.agilent.com/primerDesignProgram.jsp

### 4.5.7 Virus production

Viruses containing the desired capsid mutations were prepared through a triple plasmid transfection of HEK293T cells with the rep-cap encoding plasmid (pXX2 or pXX2-derived mutants), pSC-GFP (encodes a self-complementary GFP transgene flanked by inverted terminal repeats), and pXX6-80 (encodes adenoviral helper genes) using polyethylenimine (PEI). 48 hours post-transfection, cells were harvested and lysed by three freeze-thaw cycles. 50 U/mL of benzonase (Sigma Aldrich) was added to the cell lysate to degrade free nucleic acids and the mixture was centrifuged to remove cell debris. The supernatant was then loaded into Quick-Seal Ultra Clear 25 x 89 mm centrifuge tubes (Beckman Coulter) containing a 15-54% iodixanol step gradient. Tubes were sealed and spun at 48,000 rpm for 1 h 45 min at 18°C, and virus extracted from the 40% iodixanol layer. For the differential scanning fluorescence (DSF) assay, viruses were concentrated into gradient buffer (10 mM Tris, pH 7.6, 10 mM $MgCl_2$, 150 mM NaCl) using Amicon Ultra 100 kDa centrifugal filters (EMD Millipore).

### 4.5.8 Quantification of viral particles

Viral titers were quantified using quantitative polymerase chain reaction (qPCR). Briefly, viral capsids were denatured to release their genomes using incubation in 2M NaOH at 56 °C followed by neutralization with 2M HCl. SYBR Green Power PCR Master Mix (Thermo Fisher) along with primers against the cytomegalovirus (CMV) promoter were used to detect viral genomes. Samples were analyzed on the BioRad CFX96 qPCR machine to obtain absolute titer values against a standard curve.

### 4.5.9 Benzonase thermal stability assay

Viruses were diluted 1/10 in endo buffer (1.5 mM $MgCl_2$, 0.5 mg/mL BSA, 50 mM Tris, pH 8.0) and incubated at the indicated temperature for 30 min. Samples were then split into 20 $\mu$L treatment fractions, and treated with either 0.5 $\mu$L benzonase (1/10 dilution, 250 U/$\mu$L, Sigma), or sham buffer (50% glycerol, 50 mM Tris-HCl, 20 mM NaCl, 2 mM $MgCl_2$, pH 8.0). Samples were incubated at 37 °C for 30 min and then the benzonase was inactivated through the addition of 0.5 $\mu$L of 0.5 $\mu$M EDTA. Number of viral genomes in benzonase and sham-treated fractions were then quantified using qPCR and genome protection calculated as the ratio of genomes in the benzonase-treated fraction to the sham-treated fraction.

### 4.5.10 Differential scanning fluorescence (DSF) assay

Viruses were diluted to a concentration of $10^{12}$ viral genomes/mL in gradient buffer and 45 $\mu$L virus sample was mixed with 5 $\mu$L 50x sypro orange (Thermo Fisher). Assays were conducted in a qPCR instrument using a melt curve protocol ramping from 25 °C to 95 °C at a rate of 1 degree per min, with fluorescence reads every 0.2 degrees. Lysozyme enzyme was analyzed as a control.

### 4.5.11 Quantification of virus transduction

HEK293T cells were seeded on 48-well tissue culture-treated poly-L-lysine coated plates approximately 24 h pre-transduction. At 95% confluency, cells were transduced with virus in media with serum at 1000 multiplicity of infection (MOI). 24 h after transduction, media was changed. 48 h after transduction, cells were harvested for flow cytometry analysis on a BD FacsCantoII. Virus transduction ability was quantified using the transduction index (TI), the product of %GFP+ cells and geometric mean fluorescence intensity (gMFI). Transduction index is a linear indicator of viral transduction efficiency [147].

# Chapter 5

# Conclusions and Perspectives

The field of synthetic virology has drawn on the principles of synthetic biology to characterize viral and exogenous components as biological 'parts'. Ideally, these parts can be interchanged and combined in a modular fashion to generate variants useful for gene therapy. The work in this thesis presents two strategies for incorporating peptide and protein 'outputs' into AAV, and explores computational tools for predicting the impact of capsid genetic modification on viral formation and function. These approaches expand the synthetic virologist's toolkit for incorporating functional elements from other areas of biology into the AAV capsid, with the goal of accelerating viral engineering and broadening potential vector abilities.

Chapter 2 of this thesis describes the development of a peptide display platform based on AAV2's native stimulus-responsive display of nuclear localization signals and the PLA2 domain. Through generation of capsid protein truncation mutants, we elucidated design rules for modifying the number of peptides displayed as well as tuning activatable vs. constant display. The resulting engineered panel has variants that decorate the AAV capsid with a large number of peptides, as well as variants that activatably extrude peptides in response to temperature. In future work, this platform's potential for cellular delivery can be explored through the delivery of small bioactive peptides to intracellular and extracellular targets. A wide range of peptides are known for interacting with cellular receptors — such peptides are a straightforward, if not unique, application for this platform [181, 194]. Alternatively, we may consider peptides that exhibit extracellular bioactivity, such as peptide hormones and opioid receptor agonists [320]. To test intracellular delivery, we can focus on anti-cancer peptides that interfere with cellular metabolism and replication [320]. This peptide display platform may be further engineered for specific delivery applications. Activatable display of peptides may prove beneficial for cancer targets, due to the unique tumor microenvironment [321]. However, to fully realize this potential, the stimulus threshold of activatable display will need to be optimized. In the related parvovirus minute virus of mice (MVM), researchers have identified a capsid domain that, when modified, alters the stimulus threshold of N-terminal peptide extrusion [322]. Mutation of this region in activatable VNPs may generate a panel with different thresholds of N-terminal extrusion. This panel may then be screened for extrusion at pH and temperature conditions biologically relevant for cancer therapy. As designed, the N-terminal peptide display platform described here is not capable of gene delivery to the nucleus, as variants lack VP1 and its essential endosomal escape and nuclear localization domains. To leverage these vectors for combination peptide and gene therapy, this platform may be rendered transduction-capable with the inclusion of VP1. Initial testing has shown that incorporation of VP1 can rescue infectivity to 15% of wild-type AAV2 but optimization may further improve performance. If this peptide display platform exhibits delivery *in vivo*, it will provide a unique strategy for protecting peptides from degradation, and in

the case of activatable variants, exposure to the bloodstream. This approach may allow the development of next-generation combination therapeutics, combining the signal-transduction and targeting properties of small peptides with gene delivery.

Chapter 3 of this thesis discusses the incorporation of coiled-coil domains in the AAV capsid to facilitate modular attachment of peptides and larger proteins. This Velcro-AAV strategy avoids issues with the genetic insertion of certain motifs, such as large proteins that may disrupt capsid assembly or toxic proteins that may interfere with viral production. With this platform, both therapeutic cargo proteins and targeting motifs may be attached to the capsid. As a potential use-case, we are currently testing this platform for the delivery of CRISPR/Cas9 gene editing machinery. This mechanism of CRISPR/Cas9 delivery may reduce off-target effects due to the transience of the delivered protein and allows for co-delivery of insertion templates or other genetic motifs as viral transgenes. If this approach is successful *in vitro*, we will move forward with *in vivo* testing. Potential challenges with *in vivo* application include recognition of exposed Cas9 by the immune system and insufficient binding affinity of Cas9 to the AAV capsid for successful co-delivery. Additionally, widespread delivery of this platform may result in significant off-target effects. To address these concerns, we may further optimize this platform by coating Cas9 with lipids or polymers, covalently cross-linking the bound coiled-coil domains, and incorporating previously-developed strategies for mitigating AAV off-target delivery [211,238,323]. Successful *in vivo* editing will establish the velcro-AAV platform as a viable approach for transient targeted delivery of combined protein/DNA therapies. The coiled-coil linker design may also be useful in promoting dual-vector delivery, used in cases where the desired gene expression cassette is larger than AAV's packaging capacity. Vectors containing parts of the cassette may be linked using coiled-coil motifs to increase the chances that all components of the cassette are delivered to target cells. This velcro-AAV platform will allow synthetic virologists to consider a wider range of large proteins and structures as potential viral components.

Chapter 4 of this thesis evaluates structure-based and sequence family-based computational approaches for predicting the impact of single residue mutants on viral formation and function. We focused on the frustratometer, a characterization of protein energy landscapes based on a coarse-grained modeling approach and direct coupling analysis (DCA), a global statistical model that represents the evolutionary constraints on a sequence family. From this pilot study, we found that the frustratomer has some utility for predicting capsid formation and stability, while direct coupling analysis shows promise for predicting transduction. To validate these conclusions, these models can be screened on a larger mutational dataset spanning the structurally-resolved VP3 domain. In this subsequent analysis, we can test a wider range of mutation types to determine if our results hold for non-alanine and multi-residue mutations. We also hope to refine our computational approach to combine the best features of these models. DCA may allow researchers to identify the capsid regions most likely to tolerate mutation without perturbing viral infectivity, while AWSEM's integration of structural database information could allow the model to predict the stability of large insertional capsid mutations relevant for viral engineering. Previous integrations of these modeling frameworks have shown promise in developing a complete picture of protein interactions and energetic states. This approach applied to AAV may create an *in silico* model of the virus fitness landscape, allowing future synthetic virologists to screen their designs and iterate concepts more efficiently.

In sum, this thesis has developed a series of platforms and computational screening tools to assist virus engineers with incorporating peptides and proteins into the AAV capsid. Together, this toolkit promotes the modular insertion of biological parts into the viral capsid with predictable impacts on virus formation and function. In the spirit of synthetic biology, this work is a step towards the standardization of capsid modification — a challenge given the multifunctional roles of capsid proteins in viral infection and replication.

This approach will lead to opportunities in the integration of diverse proteins to develop biocomputing virus-based therapies.

# Bibliography


[1] L. M. Bryant, D. M. Christopher, A. R. Giles, C. Hinderer, J. L. Rodriguez, J. B. Smith, E. a. Traxler, J. Tycko, A. P. Wojno, and J. M. Wilson, "Lessons learned from the clinical development and market authorization of Glybera.," *Hum. Gene Ther.*, vol. 24, no. 2, pp. 55–64, 2013.

[2] E. Smalley, "First AAV gene therapy poised for landmark approval," *Nat. Publ. Gr.*, vol. 35, no. 11, pp. 998–999, 2017.

[3] H. Ledford, "FDA advisers back gene therapy," *Nature*, vol. 550, p. 314, 2017.

[4] R. J. Samulski and N. Muzyczka, "AAV-Mediated Gene Therapy for Research and Therapeutic Purposes," *Annu. Rev. Virol.*, vol. 1, no. 1, pp. 427–451, 2014.

[5] M. D. Weitzman and R. M. Linden, "Adeno-Associated Virus Biology," in *Adeno-Associated Virus Methods Protoc.* (R. O. Snyder and P. Moullier, eds.), vol. 807 of *Methods in Molecular Biology*, pp. 1–15, Totowa, NJ: Humana Press, 2011.

[6] F. Sonntag, K. Schmidt, and J. A. Kleinschmidt, "A Viral Assembly Factor Promotes AAV2 Capsid Formation in the Nucleolus," *Proc. Natl. Acad. Sci. USA*, vol. 107, no. 22, pp. 10220–10225, 2010.

[7] M. A. Vance, A. Mitchell, and R. J. Samulski, "AAV biology, infectivity and therapeutic use from bench to clinic," in *Gene Ther. - Princ. Challenges*, pp. 119–143, Intech, 2015.

[8] S. Misra, "Human gene therapy: a brief overview of the genetic revolution.," *J. Assoc. Physicians India*, vol. 61, no. 2, pp. 127–133, 2013.

[9] F. Sonntag, S. Bleker, B. Leuchs, R. Fischer, and J. A. Kleinschmidt, "Adeno-Associated Virus Type 2 Capsids with Externalized VP1/VP2 Trafficking Domains Are Generated Prior to Passage through the Cytoplasm and Are Maintained Until Uncoating Occurs in the Nucleus," *J. Virol.*, vol. 80, no. 22, pp. 11040–54, 2006.

[10] X. Xiao, J. Li, and R. J. Samulski, "Efficient long-term gene transfer into muscle tissue of immunocompetent mice by adeno-associated virus vector.," *J. Virol.*, vol. 70, no. 11, pp. 8098–108, 1996.

[11] A. Asokan, D. V. Schaffer, and R. J. Samulski, "The AAV vector toolkit: poised at the clinical crossroads.," *Mol. Ther.*, vol. 20, no. 4, pp. 699–708, 2012.

[12] C. Zincarelli, S. Soltys, G. Rengo, and J. E. Rabinowitz, "Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection.," *Mol. Ther.*, vol. 16, no. 6, pp. 1073–1080, 2008.

[13] S. Boutin, V. Monteilhet, P. Veron, C. Leborgne, O. Benveniste, M. F. Montus, and C. Masurier, "Prevalence of Serum IgG and Neutralizing Factors Against Adeno-Associated Virus (AAV) Types 1,

2, 5, 6, 8, and 9 in the Healthy Population: Implications for Gene Therapy Using AAV Vectors," *Hum. Gene Ther.*, vol. 21, no. 6, pp. 704–712, 2010.

[14] M. A. Kotterman and D. V. Schaffer, "Engineering Adeno-Associated Viruses for Clinical Gene Therapy," *Nat. Rev. Genet.*, vol. 15, pp. 445–51, jul 2014.

[15] C. M. Guenther, B. E. Kuypers, M. T. Lam, T. M. Robinson, J. Zhao, and J. Suh, "Synthetic Virology: Engineering Viruses for Gene Delivery," *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.*, vol. 6, no. 6, pp. 548–558, 2014.

[16] J. T. Koerber, N. Maheshri, B. K. Kaspar, and D. V. Schaffer, "Construction of diverse adeno-associated viral libraries for directed evolution of enhanced gene delivery vehicles.," *Nat. Protoc.*, vol. 1, no. 2, pp. 701–706, 2006.

[17] P. Asuri, M. A. Bartel, T. Vazin, J.-H. Jang, T. B. Wong, and D. V. Schaffer, "Directed Evolution of Adeno-associated Virus for Enhanced Gene Delivery and Gene Targeting in Human Pluripotent Stem Cells," *Mol. Ther.*, vol. 20, no. 2, pp. 329–338, 2012.

[18] D. Dalkara, L. C. Byrne, R. R. Klimczak, M. Visel, L. Yin, W. H. Merigan, J. G. Flannery, and D. V. Schaffer, "In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous.," *Sci. Transl. Med.*, vol. 5, no. 189, p. 189ra76, 2013.

[19] J. Santiago-Ortiz, D. S. Ojala, O. Westesson, J. R. Weinstein, S. Y. Wong, A. Steinsapir, S. Kumar, I. Holmes, and D. V. Schaffer, "AAV ancestral reconstruction library enables selection of broadly infectious viral variants," *Gene Ther.*, vol. 22, no. April, pp. 1–13, 2015.

[20] A. Wistuba, A. Kern, S. Weger, D. Grimm, and J. A. Kleinschmidt, "Subcellular compartmentalization of adeno-associated virus type 2 assembly.," *J. Virol.*, vol. 71, no. 2, pp. 1341–52, 1997.

[21] A. Wistuba, S. Weger, A. Kern, and J. A. Kleinschmidt, "Intermediates of Adeno-Associated Virus Type 2 Assembly: Identification of Soluble Complexes Containing Rep and Cap Proteins.," *J. Virol.*, vol. 69, no. 9, pp. 5311–9, 1995.

[22] J. Kerr, S. Cotmore, M. Bloom, R. M. Linden, and C. R. Parrish, *Parvoviruses*. Hodder Arnold, 2006.

[23] S. Steinbach, A. Wistuba, and T. Bock, "Assembly of adeno-associated virus type 2 capsids in vitro," *J. Gen. Virol.*, no. 1997, pp. 1453–1462, 1997.

[24] J. Qiu and K. E. Brown, "A 110-kDa nuclear shuttle protein, nucleolin, specifically binds to adeno-associated virus type 2 (AAV-2) capsid," *Virology*, vol. 382, pp. 373–382, 1999.

[25] J. M. Bevington, P. G. Needham, K. C. Verrill, R. F. Collaco, V. Basrur, and J. P. Trempe, "Adeno-associated virus interactions with B23/Nucleophosmin: Identification of sub-nucleolar virion regions," *Virology*, vol. 357, no. 1, pp. 102–113, 2007.

[26] J. C. Grieger, S. Snowdy, and R. J. Samulski, "Separate basic region motifs within the adeno-associated virus capsid proteins are essential for infectivity and assembly.," *J. Virol.*, vol. 80, no. 11, pp. 5199–210, 2006.

[27] L. F. Earley, Y. Kawano, K. Adachi, X.-X. Sun, M.-S. Dai, and H. Nakai, "Identification and characterization of nuclear and nucleolar localization signals in the adeno-associated virus serotype 2 assembly-activating protein," *J. Virol.*, vol. 89, no. 6, pp. 3038–3048, 2015.

[28] A. C. Maurer, S. Pacouret, A. K. Cepeda Diaz, J. Blake, E. Andres-Mateos, and L. H. Vandenberghe, "The Assembly-Activating Protein Promotes Stability and Interactions between AAV's Viral Proteins to Nucleate Capsid Assembly," *Cell Rep.*, vol. 23, no. 6, pp. 1817–1830, 2018.

[29] M. Naumer, F. Sonntag, K. Schmidt, K. Nieto, C. Panke, N. E. Davey, and R. Popa-wagner, "Properties of the adeno-associated virus assembly-activating protein," *J. Virol.*, vol. 86, no. 23, pp. 13038–13048, 2012.

[30] F. Sonntag, K. Kother, K. Schmidt, M. Weghofer, C. Raupp, K. Nieto, A. Kuck, B. Gerlach, B. Bottcher, O. J. Muller, K. Lux, M. Horer, and J. a. Kleinschmidt, "The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes," *J. Virol.*, vol. 85, no. 23, pp. 12686–12697, 2011.

[31] L. F. Earley, J. M. Powers, K. Adachi, and J. T. Baumgart, "Assembly Activating Protein Is Not an Essential Requirement for Capsid Assembly of AAV Serotypes 4, 5, and 11," *J. Virol.*, vol. 91, no. 3, pp. 1–21, 2017.

[32] D.-s. Im and N. Muzyczka, "The AAV origin binding protein Rep88 is an ATP-dependent site-specific endonuclease with DNA helicase activity," *Cell*, vol. 61, pp. 447–457, 1990.

[33] J. R. Brister and N. Muzyczka, "Mechanism of Rep-mediated adeno-associated virus origin nicking.," *J. Virol.*, vol. 74, no. 17, pp. 7762–71, 2000.

[34] T. H. Stracker, G. D. Cassell, P. Ward, Y.-m. Loo, B. van Breukelen, S. D. Carrington-Lawrence, R. K. Hamatake, P. C. van der Vliet, S. K. Weller, T. Melendy, and M. D. Weitzman, "The Rep protein of adeno-associated virus type 2 interacts with single-stranded DNA-binding proteins that enhance viral replication.," *J. Virol.*, vol. 78, no. 1, pp. 441–53, 2004.

[35] R. O. Snyder, R. J. Samulski, and N. Muzyczka, "In vitro resolution of covalently joined AAV chromosome ends," *Cell*, vol. 60, no. 1, pp. 105–113, 1990.

[36] D. Grimm, A. Kern, M. Pawlita, F. Ferrari, R. Samulski, and J. Kleinschmidt, "Titration of AAV-2 particles via a novel capsid ELISA: packaging of genomes can limit production of recombinant AAV-2.," *Gene Ther.*, vol. 6, no. 7, pp. 1322–1330, 1999.

[37] L. A. Hunter and R. J. Samulski, "Colocalization of adeno-associated virus rep and capsid proteins in the nuclei of infected cells," *J. Virol.*, vol. 66, no. 1, pp. 317–324, 1992.

[38] M. D. Weitzman, K. J. Fisher, and J. M. Wilson, "Recruitment of wild-type and recombinant adeno-associated virus into adenovirus replication centers.," *J. Virol.*, vol. 70, no. 3, pp. 1845–54, 1996.

[39] M. W. Myers and B. J. Carter, "Assembly of adeno-associated virus," *Virology*, vol. 102, no. 1, pp. 71–82, 1980.

[40] J. A. King, R. Dubielzig, D. Grimm, and J. A. Kleinschmidt, "DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids," *EMBO J.*, vol. 20, no. 12, pp. 3282–3291, 2001.

[41] K. M. Prasad and J. P. Trempe, "The adeno-associated virus Rep78 protein is covalently linked to viral DNA in a preformed virion.," *Virology*, vol. 214, no. 2, pp. 360–70, 1995.

[42] R. Dubielzig, J. A. King, S. Weger, A. Kern, and J. A. Kleinschmidt, "Adeno-associated virus type 2 protein interactions: formation of pre-encapsidation complexes," *J. Virol.*, vol. 73, no. 11, pp. 8989–8998, 1999.

[43] S. Bleker, F. Sonntag, and J. A. Kleinschmidt, "Mutational Analysis of Narrow Pores at the Fivefold Symmetry Axes of Adeno-Associated Virus Type 2 Capsids Reveals a Dual Role in Genome Packaging and Activation of Phospholipase A2 Activity," *J. Virol.*, vol. 79, no. 4, pp. 2528–40, 2005.

[44] S. Bleker, M. Pawlita, and J. A. Kleinschmidt, "Impact of Capsid Conformation and Rep-Capsid Interactions on Adeno-Associated Virus Type 2 Genome Packaging Impact of Capsid Conformation and Rep-Capsid Interactions on Adeno-Associated Virus Type 2 Genome Packaging," *J. Virol.*, vol. 80, no. 2, pp. 810–820, 2006.

[45] N. DiPrimio, A. Asokan, L. Govindasamy, M. Agbandje-McKenna, and R. J. Samulski, "Surface loop dynamics in adeno-associated virus capsid assembly," *J. Virol.*, vol. 82, no. 11, pp. 5178–5189, 2008.

[46] W. Shi, G. S. Arnold, and J. S. Bartlett, "Insertional mutagenesis of the adeno-associated virus type 2 (AAV2) capsid gene and generation of AAV2 vectors targeted to alternative cell-surface receptors.," *Hum. Gene Ther.*, vol. 12, no. 14, pp. 1697–711, 2001.

[47] S. R. Opie, K. H. Warrington, M. Agbandje-McKenna, S. Zolotukhin, and N. Muzyczka, "Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding.," *J. Virol.*, vol. 77, no. 12, pp. 6995–7006, 2003.

[48] L. M. Drouin, B. Lins, M. Janssen, A. Bennett, P. Chipman, R. McKenna, W. Chen, N. Muzyczka, G. Cardone, T. S. Baker, and M. Agbandje-McKenna, "Cryo-electron microscopy reconstruction and stability studies of wild-type and R432A variant of AAV2 reveals capsid structural stability is a major factor in genome packaging," *J. Virol.*, no. July, pp. JVI.00575–16, 2016.

[49] J. C. Grieger and R. J. Samulski, "Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps.," *J. Virol.*, vol. 79, no. 15, pp. 9933–44, 2005.

[50] N. a. Huttner, A. Girod, L. Perabo, D. Edbauer, J. a. Kleinschmidt, H. Büning, and M. Hallek, "Genetic modifications of the adeno-associated virus type 2 capsid reduce the affinity and the neutralizing effects of human serum antibodies.," *Gene Ther.*, vol. 10, no. 26, pp. 2139–47, 2003.

[51] M. a. Lochrie, G. P. Tatsuno, B. Christie, J. W. Mcdonnell, S. Zhou, G. F. Pierce, P. Colosi, and R. Surosky, "Mutations on the External Surfaces of Adeno-Associated Virus Type 2 Capsids That Affect Transduction and Neutralization Mutations on the External Surfaces of Adeno-Associated Virus Type 2 Capsids That Affect Transduction and Neutralization," *J. Virol.*, vol. 80, no. 2, pp. 821–834, 2006.

[52] C. E. Wobus, B. Hugle-Dorr, A. Girod, G. Petersen, M. Hallek, and J. A. Kleinschmidt, "Monoclonal antibodies against the adeno-associated virus type 2 (AAV-2) capsid: epitope mapping and identification of capsid domains involved in AAV-2-cell interaction and neutralization of AAV-2 infection," *J. Virol.*, vol. 74, no. 19, pp. 9281–9293, 2000.

[53] N. Kaludov, K. E. Brown, R. W. Walters, J. Zabner, and J. A. Chiorini, "Adeno-associated virus serotype 4 (AAV4) and AAV5 both require sialic acid binding for hemagglutination and efficient transduction but differ in sialic acid linkage specificity.," *J. Virol.*, vol. 75, no. 15, pp. 6884–93, 2001.

[54] Z. Wu, A. Asokan, J. C. Grieger, L. Govindasamy, M. Agbandje-McKenna, and R. J. Samulski, "Single Amino Acid Changes Can Influence Titer, Heparin Binding, and Tissue Tropism in Different Adeno-Associated Virus Serotypes," *J. Virol.*, vol. 80, no. 22, pp. 11393–11397, 2006.

[55] S. Shen, K. D. Bryant, S. M. Brown, S. H. Randell, and A. Asokan, "Terminal n-linked galactose is the primary receptor for adeno-associated virus," *J. Biol. Chem.*, vol. 286, no. 15, pp. 13532–13540, 2011.

[56] S. Pillay, N. L. Meyer, A. S. Puschnik, O. Davulcu, J. Diep, Y. Ishikawa, L. T. Jae, J. E. Wosen, C. M. Nagamine, M. S. Chapman, and J. E. Carette, "An essential receptor for adeno-associated virus infection," *Nature*, vol. 530, no. 7588, pp. 108–112, 2016.

[57] H. C. Levy, V. D. Bowman, L. Govindasamy, R. McKenna, K. Nash, K. Warrington, W. Chen, N. Muzyczka, X. Yan, T. S. Baker, and M. Agbandje-McKenna, "Heparin binding induces conformational changes in Adeno-associated virus serotype 2," *J. Struct. Biol.*, vol. 165, no. 3, pp. 146–156, 2009.

[58] C. L. Bell, B. L. Gurda, K. Van Vliet, M. Agbandje-McKenna, and J. M. Wilson, "Identification of the galactose binding domain of the adeno-associated virus serotype 9 capsid," *J. Virol.*, vol. 86, no. 13, pp. 7326–7333, 2012.

[59] L.-Y. Huang, A. Patel, R. Ng, E. B. Miller, S. Halder, R. McKenna, A. Asokan, and M. Agbandje-McKenna, "Characterization of the adeno-associated virus 1 and 6 sialic acid binding site," *J. Virol.*, vol. 90, no. March, pp. JVI.00161–16, 2016.

[60] S. Sanlioglu, P. K. Benson, J. Yang, E. M. Atkinson, T. Reynolds, and J. F. Engelhardt, "Endocytosis and nuclear trafficking of adeno-associated virus type 2 are controlled by rac1 and phosphatidylinositol-3 kinase activation.," *J. Virol.*, vol. 74, no. 19, pp. 9184–96, 2000.

[61] J. Hansen, K. Qing, H. J. Kwon, C. Mah, and A. Srivastava, "Impaired intracellular trafficking of adeno-associated virus type 2 vectors limits efficient transduction of murine fibroblasts.," *J. Virol.*, vol. 74, no. 2, pp. 992–996, 2000.

[62] B. Hauck, W. Zhao, K. High, and W. Xiao, "Intracellular viral processing, not single-stranded DNA accumulation, is crucial for recombinant adeno-associated virus transduction.," *J. Virol.*, vol. 78, no. 24, pp. 13678–13686, 2004.

[63] C. Li, Y. He, S. Nicolson, M. Hirsch, M. S. Weinberg, P. Zhang, T. Kafri, and R. J. Samulski, "Adeno-associated virus capsid antigen presentation is dependent on endosomal escape," *J. Clin. Invest.*, vol. 123, no. 3, pp. 1390–1401, 2013.

[64] W. Ding, L. N. Zhang, C. Yeaman, and J. F. Engelhardt, "rAAV2 traffics through both the late and the recycling endosomes in a dose-dependent fashion," *Mol. Ther.*, vol. 13, no. 4, pp. 671–682, 2006.

[65] Y. Liu, K.-I. Joo, and P. Wang, "Endocytic processing of adeno-associated virus type 8 vectors for transduction of target cells," *Gene Ther.*, vol. 20, no. January 2012, pp. 308–317, 2012.

[66] P.-J. Xiao and R. J. Samulski, "Cytoplasmic trafficking, endosomal escape, and perinuclear accumulation of adeno-associated virus type 2 particles are facilitated by microtubule network.," *J. Virol.*, vol. 86, no. 19, pp. 10462–73, 2012.

[67] W. Ding, L. Zhang, Z. Yan, and J. F. Engelhardt, "Intracellular trafficking of adeno-associated viral vectors.," *Gene Ther.*, vol. 12, no. 11, pp. 873–880, 2005.

[68] B. Akache, D. Grimm, X. Shen, S. Fuess, S. R. Yant, D. S. Glazer, J. Park, and M. a. Kay, "A two-hybrid screen identifies cathepsins B and L as uncoating factors for adeno-associated virus 2 and 8.," *Mol. Ther.*, vol. 15, no. 2, pp. 330–9, 2007.

[69] S. Kronenberg, B. Böttcher, C. W. von der Lieth, S. Bleker, and J. A. Kleinschmidt, "A Conformational Change in the Adeno-Associated Virus Type 2 Capsid Leads to the Exposure of Hidden VP1 N Termini," *J. Virol.*, vol. 79, no. 9, pp. 5296–5303, 2005.

[70] J. C. Grieger, J. S. Johnson, B. Gurda-Whitaker, M. Agbandje-McKenna, and R. J. Samulski, "Surface-Exposed Adeno-Associated Virus VP1-NLS Capsid Fusion Protein Rescues Infectivity of Noninfectious Wild-Type VP2/VP3 and VP3-Only Capsids but Not That of Fivefold Pore Mutant Virions," *J. Virol.*, vol. 81, no. 15, pp. 7833–7843, 2007.

[71] R. Popa-Wagner, M. Porwal, M. Kann, M. Reuss, M. Weimer, L. Florin, and J. A. Kleinschmidt, "Impact of VP1-Specific Protein Sequence Motifs on Adeno-Associated Virus Type 2 Intracellular Trafficking and Nuclear Entry," *J. Virol.*, vol. 86, no. 17, pp. 9163–9174, 2012.

[72] E. D. Horowitz, M. G. Finn, and A. Asokan, "Tyrosine Cross-Linking Reveals Interfacial Dynamics in Adeno-Associated Viral Capsids During Infection," *ACS Chem. Biol.*, vol. 7, no. 6, pp. 1059–1066, 2012.

[73] B. Venkatakrishnan, J. Yarbrough, J. Domsic, A. Bennett, B. Bothner, O. G. Kozyreva, R. J. Samulski, N. Muzyczka, R. McKenna, and M. Agbandje-McKenna, "Structure and Dynamics of Adeno-Associated Virus Serotype 1 VP1-Unique N-Terminal Domain and Its Role in Capsid Trafficking," *J. Virol.*, vol. 87, no. 9, pp. 4974–84, 2013.

[74] G. Seisenberger, "Real-Time Single-Molecule Imaging of the Infection Pathway of an Adeno-Associated Virus," *Science (80-. ).*, vol. 1929, no. 2001, 2010.

[75] S. Kelkar, B. P. De, G. Gao, J. M. Wilson, R. G. Crystal, and P. L. Leopold, "A Common Mechanism for Cytoplasmic Dynein-Dependent Microtubule Binding Shared among Adeno-Associated Virus and Adenovirus Serotypes," *J. Virol.*, vol. 80, no. 15, pp. 7781–7785, 2006.

[76] W. Zhao, L. Zhong, J. Wu, L. Chen, K. Qing, K. A. Weigel-Kelley, S. H. Larsen, W. Shou, K. H. Warrington, and A. Srivastava, "Role of cellular FKBP52 protein in intracellular trafficking of recombinant adeno-associated virus 2 vectors," *Virology*, vol. 353, no. 2, pp. 283–293, 2006.

[77] S. Hirosue, K. Senn, N. Cl??ment, M. Nonnenmacher, L. Gigout, R. M. Linden, and T. Weber, "Effect of inhibition of dynein function and microtubule-altering drugs on AAV2 transduction," *Virology*, vol. 367, no. 1, pp. 10–18, 2007.

[78] P.-J. Xiao, A. M. Mitchell, L. Huang, C. Li, and R. J. Samulski, "Disruption of microtubules post virus entry enhances adeno-associated virus vector transduction," *Hum. Gene Ther.*, vol. 27, no. 7352, p. hum.2016.008, 2016.

[79] Z. Yan, R. Zak, Y. Zhang, W. Ding, S. Godwin, K. Munson, R. Peluso, and J. F. Engelhardt, "Distinct classes of proteasome-modulating agents cooperatively augment recombinant adeno-associated virus type 2 and type 5-mediated transduction from the apical surfaces of human airway epithelia.," *J. Virol.*, vol. 78, no. 6, pp. 2863–2874, 2004.

[80] L. Denby, S. a. Nicklin, and a. H. Baker, "Adeno-associated virus (AAV)-7 and -8 poorly transduce vascular endothelial cells and are sensitive to proteasomal degradation.," *Gene Ther.*, vol. 12, no. 20, pp. 1534–8, 2005.

[81] K. Jennings, T. Miyamae, R. Traister, A. Marinov, S. Katakura, D. Sowders, B. Trapnell, J. M. Wilson, G. Gao, and R. Hirsch, "Proteasome inhibition enhances AAV-mediated transgene expression in human synoviocytes in vitro and in vivo," *Mol. Ther.*, vol. 11, no. 4, pp. 600–607, 2005.

[82] L. Zhong, W. Zhao, J. Wu, B. Li, S. Zolotukhin, L. Govindasamy, M. Agbandje-McKenna, and A. Srivastava, "A dual role of EGFR protein tyrosine kinase signaling in ubiquitination of AAV2 capsids and viral second-strand DNA synthesis.," *Mol. Ther.*, vol. 15, no. 7, pp. 1323–1330, 2007.

[83] L. Zhong, B. Li, G. Jayandharan, C. S. Mah, L. Govindasamy, M. Agbandje-McKenna, R. W. Herzog, K. A. Weigel-Van Aken, J. A. Hobbs, S. Zolotukhin, N. Muzyczka, and A. Srivastava, "Tyrosine-phosphorylation of AAV2 vectors and its consequences on viral intracellular trafficking and transgene expression," *Virology*, vol. 381, no. 2, pp. 194–202, 2008.

[84] S. C. Nicolson and R. J. Samulski, "Recombinant adeno-associated virus utilizes host cell nuclear import machinery to enter the nucleus," *J. Virol.*, vol. 88, no. 8, pp. 4132–44, 2014.

[85] J. M. Kelich, J. Ma, B. Dong, Q. Wang, M. Chin, C. M. Magura, W. Xiao, and W. Yang, "Super-resolution imaging of nuclear import of adeno-associated virus in live cells," *Mol. Ther. — Methods Clin. Dev.*, vol. 2, no. August, p. 15047, 2015.

[86] J. Hansen, "Infection of Purified Nuclei by Adeno-associated Virus 2," *Mol. Ther.*, vol. 4, no. 4, pp. 289–296, 2001.

[87] W. Xiao, K. H. W. Jr, P. Hearing, J. Hughes, N. Muzyczka, and K. H. Warrington, "Adenovirus-Facilitated Nuclear Translocation of Adeno-Associated Virus Type 2," *J. Virol.*, vol. 76, no. 22, pp. 11505–17, 2002.

[88] R. M. Tenney, C. L. Bell, and J. M. Wilson, "AAV8 capsid variable regions at the two-fold symmetry axis contribute to high liver transduction by mediating nuclear entry and capsid uncoating," *Virology*, vol. 454-455, no. 1, pp. 227–236, 2014.

[89] S. Cohen, A. R. Behzad, J. B. Carroll, and N. Panté, "Parvoviral nuclear import: Bypassing the host nuclear-transport machinery," *J. Gen. Virol.*, vol. 87, no. 11, pp. 3209–3213, 2006.

[90] C. E. Thomas, T. A. Storm, Z. Huang, and M. A. Kay, "Rapid uncoating of vector genomes is the key to efficient liver transduction with pseudotyped adeno-associated virus vectors.," *J. Virol.*, vol. 78, no. 6, pp. 3110–22, 2004.

[91] I. Sipo, H. Fechner, S. Pinkert, L. Suckau, X. Wang, S. Weger, and W. Poller, "Differential internalization and nuclear uncoating of self-complementary adeno-associated virus pseudotype vectors as determinants of cardiac cell transduction.," *Gene Ther.*, vol. 14, no. 18, pp. 1319–1329, 2007.

[92] K. Stieger, J. Schroeder, N. Provost, A. Mendes-Madeira, B. Belbellaa, G. Le Meur, M. Weber, J.-Y. Deschamps, B. Lorenz, P. Moullier, and F. Rolling, "Detection of intact rAAV particles up to 6 years after successful gene transfer in the retina of dogs and primates.," *Mol. Ther.*, vol. 17, no. 3, pp. 516–23, 2009.

[93] J. S. Bartlett, R. Wilcher, and R. J. Samulski, "Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors.," *J. Virol.*, vol. 74, pp. 2777–85, mar 2000.

[94] K. Lux, N. Goerlitz, S. Schlemminger, D. Goldnau, J. Endell, K. Leike, M. Kofler, S. Finke, M. Hallek, L. Perabo, and D. M. Kofler, "Green Fluorescent Protein-Tagged Adeno-Associated Virus Particles Allow the Study of Cytosolic and Nuclear Trafficking Green Fluorescent Protein-Tagged Adeno-Associated Virus Particles Allow the Study of Cytosolic and Nuclear Trafficking †," *J. Virol.*, vol. 79, no. 18, pp. 11776–11787, 2005.

[95] J. S. Johnson and R. J. Samulski, "Enhancement of adeno-associated virus infection by mobilizing capsids into and out of the nucleolus.," *J. Virol.*, vol. 83, no. 6, pp. 2632–2644, 2009.

[96] S. L. Murphy, A. Bhagwat, S. Edmonson, S. Zhou, and K. a. High, "High-throughput screening and biophysical interrogation of hepatotropic AAV.," *Mol. Ther.*, vol. 16, no. 12, pp. 1960–7, 2008.

[97] V. Rayaprolu, S. Kruse, R. Kant, B. Venkatakrishnan, N. Movahed, D. Brooke, B. Lins, A. Bennett, T. Potter, R. McKenna, M. Agbandje-McKenna, and B. Bothner, "Comparative Analysis of Adeno-Associated Virus Capsid Stability and Dynamics," *J. Virol.*, vol. 87, no. 24, pp. 13150–13160, 2013.

[98] E. D. Horowitz, K. S. Rahman, B. D. Bower, D. J. Dismuke, M. R. Falvo, J. D. Griffith, S. C. Harvey, and A. Asokan, "Biophysical and ultrastructural characterization of adeno-associated virus capsid uncoating and genome release.," *J. Virol.*, vol. 87, no. 6, pp. 2994–3002, 2013.

[99] L. Zhong, X. Zhou, Y. Li, K. Qing, X. Xiao, R. J. Samulski, and A. Srivastava, "Single-polarity recombinant adeno-associated virus 2 vector-mediated transgene expression in vitro and in vivo: mechanism of transduction.," *Mol. Ther.*, vol. 16, no. 2, pp. 290–5, 2008.

[100] X. Zhou, X. Zeng, Z. Fan, C. Li, T. McCown, R. J. Samulski, and X. Xiao, "Adeno-associated virus of a single-polarity DNA genome is capable of transduction in vivo.," *Mol. Ther.*, vol. 16, no. 3, pp. 494–9, 2008.

[101] K. J. Fisher, G. P. Gao, M. D. Weitzman, R. DeMatteo, J. F. Burda, and J. M. Wilson, "Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis.," *J. Virol.*, vol. 70, no. 1, pp. 520–532, 1996.

[102] F. K. Ferrari, T. Samulski, T. Shenk, and R. J. Samulski, "Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors.," *J. Virol.*, vol. 70, no. 5, pp. 3227–34, 1996.

[103] K. Qing, X. S. Wang, D. M. Kube, S. Ponnazhagan, A. Bajpai, and A. Srivastava, "Role of tyrosine phosphorylation of a cellular protein in adeno-associated virus 2-mediated transgene expression," *Proc. Natl. Acad. Sci. USA*, vol. 94, no. 20, pp. 10879–10884, 1997.

[104] K. Qing, W. Li, L. Zhong, M. Tan, J. Hansen, K. A. Weigel-Kelley, L. Chen, M. C. Yoder, and A. Srivastava, "Adeno-associated virus type 2-mediated gene transfer: role of cellular T-cell protein tyrosine phosphatase in transgene expression in established cell lines in vitro and transgenic mice in vivo.," *J. Virol.*, vol. 77, no. 4, pp. 2741–2746, 2003.

[105] J. Jurvansuu, K. Raj, A. Stasiak, and P. Beard, "Viral transport of DNA damage that mimics a stalled replication fork.," *J. Virol.*, vol. 79, no. 1, pp. 569–580, 2005.

[106] R. a. Schwartz, J. A. Palacios, G. D. Cassell, S. Adam, M. Giacca, and M. D. Weitzman, "The Mre11/Rad50/Nbs1 complex limits adeno-associated virus transduction and replication.," *J. Virol.*, vol. 81, no. 23, pp. 12936–45, 2007.

[107] T. Cervelli, J. A. Palacios, L. Zentilin, M. Mano, R. a. Schwartz, M. D. Weitzman, and M. Giacca, "Processing of recombinant AAV genomes occurs in specific nuclear structures that overlap with foci of DNA-damage-response proteins.," *J. Cell Sci.*, vol. 121, no. Pt 3, pp. 349–57, 2008.

[108] M. Salganik, F. Aydemir, H.-J. Nam, R. McKenna, M. Agbandje-McKenna, and N. Muzyczka, "Adeno-associated virus capsid proteins may play a role in transcription and second-strand synthesis of recombinant genomes.," *J. Virol.*, vol. 88, pp. 1071–9, jan 2014.

[109] F. Aydemir, M. Salganik, J. Resztak, J. Singh, A. Bennett, M. Agbandje-Mckenna, and N. Muzyczka, "Mutants at the 2-fold interface of AAV2 structural proteins suggest a role in viral transcription for AAV capsids," *J. Virol.*, no. June, 2016.

[110] Y. Yue and D. Duan, "Double strand interaction is the predominant pathway for intermolecular recombination of adeno-associated viral genomes," *Virology*, vol. 313, no. 1, pp. 1–7, 2003.

[111] H. Nakai, T. A. Storm, S. Fuess, and M. A. Kay, "Pathways of removal of free DNA vector ends in normal and DNA-PKcs-deficient SCID mouse hepatocytes transduced with rAAV vectors," *Hum. Gene Ther.*, vol. 14, no. 9, pp. 871–881, 2003.

[112] S. Ganta, H. Devalapally, A. Shahiwala, and M. Amiji, "A review of stimuli-responsive nanocarriers for drug and gene delivery," *J. Controlled Release*, vol. 126, no. 3, pp. 187–204, 2008.

[113] Z. Kis, S. Pereira, T. Homma, R. M. Pedrigi, and R. Krams, "Mammalian synthetic biology : emerging medical applications Headline review Synthetic biosystems," *R. Soc. Interface*, vol. 12, no. 20141000, pp. 1–18, 2015.

[114] H. Ye and M. Fussenegger, "Synthetic therapeutic gene circuits in mammalian cells," *FEBS Lett.*, vol. 588, no. 15, pp. 2537–2544, 2014.

[115] A. S. Khalil and J. J. Collins, "Synthetic biology: applications come of age.," *Nat. Rev. Genet.*, vol. 11, no. 5, pp. 367–379, 2010.

[116] F. Lienert, J. J. Lohmueller, A. Garg, and P. a. Silver, "Synthetic biology in mammalian cells: next generation research tools and therapeutics.," *Nat. Rev. Mol. Cell Biol.*, vol. 15, no. 2, pp. 95–107, 2014.

[117] J. Y. Kim and J. S. Lee, "Synthesis and thermally reversible assembly of DNA - Gold nanoparticle cluster conjugates," *Nano Lett.*, vol. 9, no. 12, pp. 4564–4569, 2009.

[118] X. Qian, J. Li, and S. Nie, "Stimuli-responsive SERS nanoparticles: conformational control of plasmonic coupling and surface Raman enhancement.," *J. Am. Chem. Soc.*, vol. 131, no. 22, pp. 7540–1, 2009.

[119] X. Qian, X. Zhou, and S. Nie, "Surface-enhanced raman nanoparticle beacons based on bioconjugated gold nanocrystals and long range plasmonic coupling," *J. Am. Chem. Soc.*, vol. 130, no. 45, pp. 14934–14935, 2008.

[120] J. M. Perez, L. Josephson, T. O'Loughlin, D. Högemann, and R. Weissleder, "Magnetic Relaxation Switches Capable of Sensing Molecular Interactions," *Nat. Biotechnol.*, vol. 20, no. 8, pp. 816–820, 2002.

[121] C. Sonnichsen, B. M. Reinhard, J. Liphard, and A. P. Alivisatos, "A molecular ruler based on plasmon coupling of single gold and silver nanoparticles," *Nat. Biotechnol.*, vol. 23, no. 6, pp. 741–745, 2005.

[122] J. Yang, C. Dong, Y. Dong, S. Liu, L. Pan, and C. Zhang, "Logic Nanoparticle Beacon Triggered by the Binding-Induced Effect of Multiple Inputs," *Appl. Mater. Interfaces*, pp. 4–10, 2014.

[123] C. Zhang, J. Yang, and J. Xu, "Molecular logic computing model based on self-assembly of DNA nanoparticles," *Chinese Sci. Bull.*, vol. 56, no. 33, pp. 3566–3571, 2011.

[124] X. Xu, J. Zhang, F. Yang, and X. Yang, "Colorimetric logic gates for small molecules using split/integrated aptamers and unmodified gold nanoparticles.," *Chem. Commun. (Camb).*, vol. 47, no. 33, pp. 9435–9437, 2011.

[125] D. Liu, W. Chen, K. Sun, K. Deng, W. Zhang, Z. Wang, and X. Jiang, "Resettable, Multi-Readout Logic Gates Based on Controllably Reversible Aggregation of Gold Nanoparticles," *Angew. Chemie - Int. Ed.*, vol. 50, no. 18, pp. 4103–4107, 2011.

[126] A. P. Majewski and A. Schallon, "Dual-Responsive Magnetic Core-Shell Nanoparticles for Nonviral Gene Delivery and Cell Separation," *Biomacromolecules*, 2012.

[127] R. Visse and H. Nagase, "Matrix metalloproteinases and tissue inhibitors of metalloproteinases: Structure, function, and biochemistry," *Circ. Res.*, vol. 92, no. 8, pp. 827–839, 2003.

[128] H. Birkedal-Hansen, W. G. I. Moore, M. K. Bodden, L. J. Windsor, B. Birkedal-Hansen, A. Decarlo, and J. A. Engler, "Matrix Metalloproteinases : A Review *," *Crit. Rev. Oral Biol. Med.*, vol. 4, no. 2, pp. 197–250, 1993.

[129] G. Von Maltzahn, T. J. Harris, J. H. Park, D. H. Min, A. J. Schmidt, M. J. Sailor, and S. N. Bhatia, "Nanoparticle self-assembly gated by logical proteolytic triggers," *J. Am. Chem. Soc.*, vol. 129, no. 19, pp. 6064–6065, 2007.

[130] M. P. Nikitin, V. O. Shipunova, S. M. Deyev, and P. I. Nikitin, "Biocomputing based on particle disassembly," *Nat. Nanotechnol.*, vol. 9, pp. 716–722, aug 2014.

[131] S. J. Tseng, I. M. Kempson, S. F. Peng, B. H. Ke, H. H. Chen, P. F. Chen, and Y. Hwu, "Environment Acidity Triggers Release of Recombinant Adeno-Associated Virus Serotype 2 from a Tunable Matrix," *J. Controlled Release*, vol. 170, no. 2, pp. 252–258, 2013.

[132] Y. Kotsuchibashi, M. Ebara, a. S. Hoffman, R. Narain, and T. Aoyagi, "Temperature-responsive mixed core nanoparticle properties determined by the composition of statistical and block copolymers in the core," *Polym. Chem.*, pp. 1693–1697, 2015.

[133] R. Cheng, F. Meng, C. Deng, H. A. Klok, and Z. Zhong, "Dual and multi-stimuli responsive polymeric nanoparticles for programmed site-specific drug delivery," *Biomaterials*, vol. 34, no. 14, pp. 3647–3657, 2013.

[134] E. a. Mahmoud, J. Sankaranarayanan, J. M. Morachis, G. Kim, and A. Almutairi, "Inflammation responsive logic gate nanoparticles for the delivery of proteins," *Bioconjug. Chem.*, vol. 22, no. 7, pp. 1416–1421, 2011.

[135] L. Bai, X.-h. Wang, F. Song, X.-l. Wang, and Y.-z. Wang, ""AND" Logic Gate Regulated pH and Reduction Dual-Responsive Prodrug Nanoparticles for Efficient Intracellular Anticancer Drug Delivery," *Chem. Communcations*, vol. 51, no. 1, pp. 93–96, 2015.

[136] Y. Shen, X. Ma, B. Zhang, Z. Zhou, Q. Sun, E. Jin, M. Sui, J. Tang, J. Wang, and M. Fan, "Degradable dual pH- and temperature-responsive photoluminescent dendrimers," *Chem. - A Eur. J.*, vol. 17, no. 19, pp. 5319–5326, 2011.

[137] Y. C. Chen, L. C. Liao, P. L. Lu, C. L. Lo, H. C. Tsai, C. Y. Huang, K. C. Wei, T. C. Yen, and G. H. Hsiue, "The Accumulation of Dual pH and Temperature Responsive Micelles in Tumors," *Biomaterials*, vol. 33, no. 18, pp. 4576–4588, 2012.

[138] Q. Huang, T. Liu, C. Bao, Q. Lin, M. Ma, and L. Zhu, "Light and Reductive Dual Stimuli-Responsive PEI Nanoparticles: "AND" Logic Response and Controllable Release," *J. Mater. Chem. B*, vol. 2, no. 21, pp. 3333–3339, 2014.

[139] M. Oishi, T. Nakamura, Y. Jinji, K. Matsuishi, and Y. Nagasaki, "Multi-stimuli-triggered release of charged dye from smart PEGylated nanogels containing gold nanoparticles to regulate fluorescence signals," *J. Mater. Chem.*, vol. 19, no. 33, p. 5909, 2009.

[140] F. Pu, J. Ren, and X. Qu, ""Plug and Play" Logic Gates Based on Fluorescence Switching Regulated by Self-Assembly of Nucleotide and Lanthanide Ions," *ACS Appl. Mater. Interfaces*, vol. 6, no. 12, pp. 9557–9562, 2014.

[141] F. Pu, E. Ju, J. Ren, and X. Qu, "Multiconfigurable logic gates based on fluorescence switching in adaptive coordination polymer nanoparticles," *Adv. Mater.*, vol. 26, no. 7, pp. 1111–1117, 2014.

[142] M. Motornov, J. Zhou, M. Pita, V. Gopishetty, I. Tokarev, E. Katz, and S. Minko, ""Chemical transformers" from nanoparticle ensembles operated with logic," *Nano Lett.*, vol. 8, no. 9, pp. 2993–2997, 2008.

[143] M. Pita, M. Krämer, J. Zhou, A. Poghossian, M. J. Schöning, V. M. Fernández, and E. Katz, "Optoelectronic properties of nanostructured ensembles controlled by biomolecular logic systems," *ACS Nano*, vol. 2, no. 10, pp. 2160–2166, 2008.

[144] J. Du, S. Yin, L. Jiang, B. Ma, and X. Chen, "A Colorimetric Logic Gate Based on Free Gold Nanoparticles and Coordination Strategy between Melamine and Mercury Ions," *Chem. Commun.*, pp. 4196–4198, 2013.

[145] S. M. Douglas, I. Bachelet, and G. Church, "A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads," *Science (80-. ).*, vol. 335, pp. 831–834, 2012.

[146] B. Yang, X.-B. Zhang, L.-P. Kang, Z.-M. Huang, G.-L. Shen, R.-Q. Yu, and W. Tan, "Intelligent layered nanoflare: "lab-on-a-nanoparticle" for multiple DNA logic gate operations and efficient intracellular delivery," *Nanoscale*, vol. 6, no. 15, p. 8990, 2014.

[147] J. Judd, M. L. Ho, A. Tiwari, E. J. Gomez, C. Dempsey, K. Van Vliet, O. A. Igoshin, J. J. Silberg, M. Agbandje-Mckenna, and J. Suh, "Tunable Protease-Activatable Virus Nanonodes," *ACS Nano*, vol. 8, pp. 4740–6, may 2014.

[148] D. Vartak and R. Gemeinhart, "Matrix metalloproteases: Underutilized targets for drug delivery," *J. Drug Target.*, vol. 15, no. 1, pp. 1–20, 2007.

[149] S. Angelos, Y. W. Yang, N. M. Khashab, J. F. Stoddart, and J. I. Zink, "Dual-controlled nanoparticles exhibiting and logic," *J. Am. Chem. Soc.*, vol. 131, no. 32, pp. 11344–11346, 2009.

[150] J.-W. Oh, W.-J. Chung, K. Heo, H.-E. Jin, B. Y. Lee, E. Wang, C. Zueger, W. Wong, J. Meyer, C. Kim, S.-Y. Lee, W.-G. Kim, M. Zemla, M. Auer, A. Hexemer, and S.-W. Lee, "Biomimetic virus-based colourimetric sensors.," *Nat. Commun.*, vol. 5, p. 3043, 2014.

[151] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, and K. Robins, "Engineering the third wave of biocatalysis," *Nature*, vol. 485, no. 7397, pp. 185–194, 2012.

[152] D. Hilvert, "Design of protein catalysts.," *Annu. Rev. Biochem.*, vol. 82, pp. 447–70, 2013.

[153] C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard, and D. Baker, "Computational design of ligand-binding proteins with high affinity and selectivity.," *Nature*, vol. 501, no. 7466, pp. 212–6, 2013.

[154] H. Kries, R. Blomberg, and D. Hilvert, "De novo enzymes by computational design," *Curr. Opin. Chem. Biol.*, vol. 17, no. 2, pp. 221–228, 2013.

[155] M. P. Frushicheva, M. J. L. Mills, P. Schopf, M. K. Singh, R. B. Prasad, and A. Warshel, "Computer aided enzyme design and catalytic concepts," *Curr. Opin. Chem. Biol.*, vol. 21, pp. 56–62, 2014.

[156] J. Damborsky and J. Brezovsky, "Computational tools for designing and engineering enzymes," *Curr. Opin. Chem. Biol.*, vol. 19, no. 1, pp. 8–16, 2014.

[157] M. Goldsmith and D. S. Tawfik, "Directed enzyme evolution: Beyond the low-hanging fruit," *Curr. Opin. Struct. Biol.*, vol. 22, no. 4, pp. 406–412, 2012.

[158] M. J. Harms and J. W. Thornton, "Evolutionary biochemistry: revealing the historical and physical causes of protein properties.," *Nat. Rev. Genet.*, vol. 14, no. 8, pp. 559–71, 2013.

[159] M. S. Packer and D. R. Liu, "Methods for the directed evolution of proteins.," *Nat. Rev. Genet.*, vol. 16, no. 7, pp. 379–394, 2015.

[160] A. Currin, N. Swainston, P. J. Day, and D. B. Kell, "Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently.," *Chem. Soc. Rev.*, vol. 44, no. 5, pp. 1172–239, 2015.

[161] A. M. Wen and N. F. Steinmetz, "Design of Virus-Based Nanomaterials for Medicine, Biotechnology, and Energy," *Chem. Soc. Rev.*, vol. 45, no. 15, pp. 4074–4126, 2016.

[162] A. C. Evans, N. N. Thadani, and J. Suh, "Biocomputing Nanoplatforms as Therapeutics and Diagnostics," *J. Controlled Release*, vol. 240, pp. 387–393, 2015.

[163] A. Hajitou, M. Trepel, C. E. Lilley, S. Soghomonyan, M. M. Alauddin, F. C. Marini, B. H. Restel, M. G. Ozawa, C. A. Moya, R. Rangel, Y. Sun, K. Zaoui, M. Schmidt, C. von Kalle, M. D. Weitzman, J. G. Gelovani, R. Pasqualini, and W. Arap, "A Hybrid Vector for Ligand-Directed Tumor Targeting and Molecular Imaging," *Cell*, vol. 125, pp. 385–398, 2006.

[164] K. White, H. Büning, A. Kritz, H. Janicki, J. McVey, L. Perabo, G. Murphy, M. Odenthal, L. M. Work, M. Hallek, S. A. Nicklin, and A. H. Baker, "Engineering Adeno-Associated Virus 2 Vectors for Targeted Gene Delivery to Atherosclerotic Lesions," *Gene Ther.*, vol. 15, pp. 443–51, mar 2008.

[165] E. J. Gomez, K. Gerhardt, J. Judd, J. J. Tabor, and J. Suh, "Light-Activated Nuclear Translocation of Adeno-Associated Virus Nanoparticles Using Phytochrome B for Enhanced, Tunable, and Spatially Programmable Gene Delivery," *ACS Nano*, vol. 10, no. 1, pp. 225–237, 2016.

[166] E. Kim, J.-S. Oh, I.-S. Ahn, K. I. Park, and J.-H. Jang, "Magnetically Enhanced Adeno-Associated Viral Vector Delivery for Human Neural Stem Cell Infection," *Biomaterials*, vol. 32, no. 33, pp. 8654–8662, 2011.

[167] C. A. Hong, S. K. Cho, J. A. Edson, J. Kim, D. Ingato, B. Pham, A. Chuang, D. A. Fruman, and Y. J. Kwon, "Viral/Nonviral Chimeric Nanoparticles to Synergistically Suppress Leukemia Proliferation via Simultaneous Gene Transduction and Silencing," *ACS Nano*, vol. 10, no. 9, pp. 8705–8714, 2016.

[168] S. Suikkanen, M. Antila, A. Jaatinen, M. Vihinen-Ranta, and M. Vuento, "Release of Canine Parvovirus From Endocytic Vesicles," *Virology*, vol. 316, no. 2, pp. 267–280, 2003.

[169] B. Kaufmann, P. R. Chipman, V. A. Kostyuchenko, S. Modrow, and M. G. Rossmann, "Visualization of the Externalized VP2 N Termini of Infectious Human Parvovirus B19," *J. Virol.*, vol. 82, no. 15, pp. 7306–12, 2008.

[170] S. F. Cotmore, S. Hafenstein, and P. Tattersall, "Depletion of Virion-Associated Divalent Cations Induces Parvovirus Minute Virus of Mice to Eject Its Genome in a 3'-to-5' Direction from an Otherwise Intact Viral Particle," *J. Virol.*, vol. 84, no. 4, pp. 1945–56, 2010.

[171] S. Kailasan, S. Halder, B. Gurda, H. Bladek, P. R. Chipman, R. McKenna, K. Brown, and M. Agbandje-McKenna, "Structure of an Enteric Pathogen, Bovine Parvovirus," *J. Virol.*, vol. 89, no. 5, pp. 2603–2614, 2015.

[172] M. A. Musick, K. I. McConnell, J. K. Lue, F. Wei, C. Chen, and J. Suh, "Reprogramming Virus Nanoparticles to Bind Metal Ions upon Activation With Heat," *Biomacromolecules*, vol. 12, pp. 2153–2158, 2011.

[173] K. H. Warrington Jr, O. S. Gorbatyuk, J. K. Harrison, S. R. Opie, S. Zolotukhin, and N. Muzyczka, "Adeno-Associated Virus Type 2 VP2 Capsid Protein Is Nonessential and Can Tolerate Large Peptide Insertions at Its N Terminus," *J. Virol.*, vol. 78, no. 12, pp. 6595–6609, 2004.

[174] M. P. Williamson, "The Structure and Function of Proline-Rich Regions in Proteins," *Biochem. J.*, vol. 297, no. 2, pp. 249–260, 1994.

[175] E. Ruoslahti, S. N. Bhatia, and M. J. Sailor, "Targeting of Drugs and Nanoparticles to Tumors," *J. Cell Biol.*, vol. 188, no. 6, pp. 759–768, 2010.

[176] D. Wang, W. Li, H. Zhang, Q. Mao, and H. Xia, "A Targeting Peptide Improves Adenovirus-Mediated Transduction of a Glioblastoma Cell Line," *Oncol. Rep.*, vol. 31, no. 5, pp. 2093–2098, 2014.

[177] Y. Liu, Y. J. Kim, M. Ji, J. Fang, N. Siriwon, L. I. Zhang, and P. Wang, "Enhancing Gene Delivery of Adeno-Associated Viruses by Cell-Permeable Peptides," *Mol. Ther. - Methods Clin. Dev.*, vol. 1, no. December 2013, pp. 1–13, 2014.

[178] W. T. Hermens, O. Ter Brake, P. A. Dijkhuizen, M. A. Sonnemans, D. Grimm, J. A. Kleinschmidt, and J. Verhaagen, "Purification of Recombinant Adeno-Associated Virus by Iodixanol Gradient Ultracentrifugation Allows Rapid and Reproducible Preparation of Vector Stocks for Gene Transfer in the Nervous System," *Hum. Gene Ther.*, vol. 10, no. 11, pp. 1885–1991, 1999.

[179] M. J. Brun, E. J. Gomez, and J. Suh, "Stimulus-responsive viral vectors for controlled delivery of therapeutics," *J. Controlled Release*, vol. 267, no. August, pp. 80–89, 2017.

[180] P. Wu, W. Xiao, T. Conlon, J. Hughes, M. Agbandje-McKenna, T. Ferkol, T. Flotte, and N. Muzyczka, "Mutational analysis of the adeno-associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism.," *J. Virol.*, vol. 74, no. 18, pp. 8635–8647, 2000.

[181] M. Grifman, M. Trepel, P. Speece, L. B. Gilbert, W. Arap, R. Pasqualini, and M. D. Weitzman, "Incorporation of tumor-targeting peptides into recombinant adeno-associated virus capsids," *Mol. Ther.*, vol. 3, no. 6, pp. 964–975, 2001.

[182] L. Gigout, P. Rebollo, N. Clement, K. H. Warrington, N. Muzyczka, R. M. Linden, and T. Weber, "Altering AAV tropism with mosaic viral capsids," *Mol. Ther.*, vol. 11, no. 6, pp. 856–865, 2005.

[183] X. Shen, T. Storm, and M. A. Kay, "Characterization of the relationship of AAV capsid domain swapping to liver transduction efficiency.," *Mol. Ther.*, vol. 15, no. 11, pp. 1955–1962, 2007.

[184] W. Li, A. Asokan, Z. Wu, T. Van Dyke, N. DiPrimio, J. S. Johnson, L. Govindaswamy, M. Agbandje-McKenna, S. Leichtle, D. E. Redmond, T. J. McCown, K. B. Petermann, N. E. Sharpless, and R. J. Samulski, "Engineering and selection of shuffled AAV genomes: a new strategy for producing targeted biological nanoparticles.," *Mol. Ther.*, vol. 16, no. 7, pp. 1252–1260, 2008.

[185] P. Ward and C. E. Walsh, "Chimeric AAV Cap sequences alter gene transduction," *Virology*, vol. 386, no. 2, pp. 237–248, 2009.

[186] H. Büning, A. Huber, L. Zhang, N. Meumann, and U. Hacker, "Engineering the AAV capsid to optimize vector-host-interactions," *Curr. Opin. Pharmacol.*, vol. 24, no. August, pp. 94–104, 2015.

[187] J. Körbelin, T. Sieber, S. Michelfelder, L. Lunding, E. Spies, A. Hunger, M. Alawi, K. Rapti, D. Indenbirken, O. J. Müller, R. Pasqualini, W. Arap, J. A. Kleinschmidt, and M. Trepel, "Pulmonary Targeting of Adeno-associated Viral Vectors by Next-generation Sequencing-guided Screening of Random Capsid Displayed Peptide Libraries.," *Mol. Ther.*, vol. 24, no. 6, pp. 1050–1061, 2016.

[188] F. Wei, K. I. McConnell, T.-K. Yu, and J. Suh, "Conjugation of paclitaxel on adeno-associated virus (AAV) nanoparticles for co-delivery of genes and drugs.," *Eur. J. Pharm. Sci.*, vol. 46, pp. 167–72, jun 2012.

[189] N. N. Thadani, C. Dempsey, J. Zhao, S. M. Vasquez, and J. Suh, "Reprogramming the Activatable Peptide Display Function of Adeno-Associated Virus Nanoparticles," *ACS Nano*, p. acsnano.7b07804, 2017.

[190] M. Chen, X. Lin, W. Lu, N. P. Schafer, J. N. Onuchic, and P. G. Wolynes, "Template-Guided Protein Structure Prediction and Refinement Using Optimized Folding Landscape Force Fields," *J. Chem. Theory Comput.*, vol. 14, no. 11, pp. 6102–6116, 2018.

[191] J. S. Bartlett, J. Kleinschmidt, R. C. Boucher, and R. J. Samulski, "Targeted adeno-associated virus vector transduction of nonpermissive cells mediated by a bispecific f(ab'$\gamma$)2antibody," *Nat. Biotechnol.*, vol. 17, no. 2, pp. 181–186, 1999.

[192] S. Ponnazhagan, G. Mahendra, S. Kumar, J. A. Thompson, and M. Castillas, "Conjugate-based targeting of recombinant adeno-associated virus type 2 vectors by using avidin-linked ligands.," *J. Virol.*, vol. 76, no. 24, pp. 12900–7, 2002.

[193] G. S. Arnold, A. K. Sasser, M. D. Stachler, and J. S. Bartlett, "Metabolic Biotinylation Provides a Unique Platform for the Purification and Targeting of Multiple AAV Vector Serotypes," *Mol. Ther.*, vol. 14, no. 1, pp. 97–106, 2006.

[194] M. D. Stachler, I. Chen, A. Y. Ting, and J. S. Bartlett, "Site-specific modification of AAV vector particles with biophysical probes and targeting ligands using biotin ligase," *Mol. Ther.*, vol. 16, no. 8, pp. 1467–1473, 2008.

[195] R. E. Kelemen, R. Mukherjee, X. Cao, S. B. Erickson, Y. Zheng, and A. Chatterjee, "A Precise Chemical Strategy To Alter the Receptor Specificity of the Adeno-Associated Virus," *Angew. Chemie Int. Ed.*, pp. 1–6, 2016.

[196] D. Katrekar, A. M. Moreno, G. Chen, A. Worlikar, and P. Mali, "Oligonucleotide conjugated multifunctional adeno-associated viruses," *Sci. Rep.*, vol. 8, no. 1, pp. 1–8, 2018.

[197] C. Zhang, X. Zhou, T. Yao, Z. Tian, and D. Zhou, "Precision Fluorescent Labeling of an Adeno-Associated Virus Vector to Monitor the Viral Infection Pathway," *Biotechnol. J.*, vol. 1700374, p. 1700374, 2018.

[198] C. D. Hein, X. M. Liu, and D. Wang, "Click chemistry, a powerful tool for pharmaceutical sciences," *Pharm. Res.*, vol. 25, no. 10, pp. 2216–2230, 2008.

[199] S. C. Yadav, A. Kumari, and R. Yadav, "Development of peptide and protein nanotherapeutics by nanoencapsulation and nanobioconjugation," *Peptides*, vol. 32, no. 1, pp. 173–187, 2011.

[200] A. Lupas, "Coiled coils: New structures and new functions," *Trends Biochem. Sci.*, vol. 21, no. 10, pp. 375–382, 1996.

[201] A. N. Lupas and J. Bassler, "Coiled Coils – A Model System for the 21st Century," *Trends Biochem. Sci.*, vol. 42, no. 2, pp. 130–140, 2017.

[202] J. M. Mason and K. M. Arndt, "Coiled coil domains: Stability, specificity, and biological implications," *ChemBioChem*, vol. 5, no. 2, pp. 170–176, 2004.

[203] D. K. Woolfson, "The Design of Coiled-Coil Structures and Assemblies," *Adv. Protein Chem.*, vol. 70, no. 04, pp. 79–112, 2005.

[204] E. K. O'Shea, R. Rutkowski, and P. S. Kim, "Evidence That the Leucine Zipper Is a Coiled Coil," *Science (80-. ).*, vol. 243, no. 4890, pp. 538–542, 1989.

[205] W. H. Landschulz, P. F. Johnson, and S. L. Mcknight, "The Leucine Zipper: A Hypothetical Structure Common to a New Class of DNA Binding Proteins," *Science (80-. ).*, vol. 240, no. 4860, pp. 1759–1764, 1988.

[206] B. Tripet, L. Yu, D. L. Bautista, W. Y. Wong, R. T. Irvin, and R. S. Hodges, "Engineering a de novo-designed coiled-coil heterodimerization domain for the rapid detection, purification and characterization of recombinantly expressed peptides and proteins," *Protein Eng.*, vol. 9, no. 11, pp. 1029–1042, 1996.

[207] J. R. Litowski and R. S. Hodges, "Designing heterodimeric two-stranded $\alpha$-helical coiled-coils. Effects of hydrophobicity and $\alpha$-helical propensity on protein folding, stability, and specificity," *J. Biol. Chem.*, vol. 277, no. 40, pp. 37272–37279, 2002.

[208] H. Chao, D. L. Bautista, J. Litowski, R. T. Irvin, and R. S. Hodges, "Use of a heterodimeric coiled-coil system for biosensor application and affinity purification," *J. Chromatogr. B Biomed. Appl.*, vol. 715, no. 1, pp. 307–329, 1998.

[209] B. Apostolovic and H. A. Klok, "pH-sensitivity of the E3/K3 heterodimeric coiled coil," *Biomacromolecules*, vol. 9, no. 11, pp. 3173–3180, 2008.

[210] J. Fernandez-Rodriguez and T. C. Marlovits, "Induced heterodimerization and purification of two target proteins by a synthetic coiled-coil tag," *Protein Sci.*, vol. 21, no. 4, pp. 511–519, 2012.

[211] J. Wang, Y. Yu, and J. Xia, "Short peptide tag for covalent protein labeling based on coiled coils," *Bioconjug. Chem.*, vol. 25, no. 1, pp. 178–187, 2014.

[212] C. Minelli, J. X. Liew, M. Muthu, and H. Andresen, "Coiled coil peptide-functionalized surfaces for reversible molecular binding," *Soft Matter*, vol. 9, no. 20, p. 5119, 2013.

[213] J. M. Fletcher, R. L. Harniman, F. R. H. Barnes, A. L. Boyle, A. Collins, J. Mantell, T. H. Sharp, M. Antognozzi, P. J. Booth, N. Linden, M. J. Miles, R. B. Sessions, P. Verkade, and D. N. Woolfson, "Coiled-Coil Peptide Modules," *Science (80-. ).*, vol. 340, no. May, pp. 595–599, 2013.

[214] Y. Assal, Y. Mizuguchi, M. Mie, and E. Kobatake, "Growth Factor Tethering to Protein Nanoparticles via Coiled-Coil Formation for Targeted Drug Delivery," *Bioconjug. Chem.*, vol. 26, no. 8, pp. 1672–1677, 2015.

[215] B. Apostolovic, S. P. Deacon, R. Duncan, and H. A. Klok, "Hybrid polymer therapeutics incorporating bioresponsive, coiled coil peptide linkers," *Biomacromolecules*, vol. 11, no. 5, pp. 1187–1195, 2010.

[216] B. Apostolovic, S. P. Deacon, R. Duncan, and H. A. Klok, "Cell uptake and trafficking behavior of non-covalent, coiled-coil based polymer-drug conjugates," *Macromol. Rapid Commun.*, vol. 32, no. 1, pp. 11–18, 2011.

[217] M. Pechar and R. Pola, "The coiled coil motif in polymer drug delivery systems," *Biotechnol. Adv.*, vol. 31, no. 1, pp. 90–96, 2013.

[218] M. M. Stevens, N. T. Flynn, C. Wang, D. A. Tirrell, and R. Langer, "Coiled-coil peptide-based assembly of gold nanoparticles," *Adv. Mater.*, vol. 16, no. 11, pp. 915–918, 2004.

[219] S. C. Wagner, M. Roskamp, H. Cölfen, C. Böttcher, S. Schlecht, and B. Koksch, "Switchable electrostatic interactions between gold nanoparticles and coiled coilpeptides direct colloid assembly," *Org. Biomol. Chem.*, vol. 7, no. 1, pp. 46–51, 2009.

[220] G. Martelli, H. R. Zope, M. Bròvia Capell, and A. Kros, "Coiled-coil peptide motifs as thermoresponsive valves for mesoporous silica nanoparticles," *Chem. Commun.*, vol. 49, no. 85, p. 9932, 2013.

[221] L. Ding, Y. Jiang, J. Zhang, H. A. Klok, and Z. Zhong, "PH-Sensitive Coiled-Coil Peptide-Cross-Linked Hyaluronic Acid Nanogels: Synthesis and Targeted Intracellular Protein Delivery to CD44 Positive Cancer Cells," *Biomacromolecules*, vol. 19, no. 2, pp. 555–562, 2018.

[222] J. Yang, Y. Shimada, R. C. Olsthoorn, B. E. Snaar-Jagalska, H. P. Spaink, and A. Kros, "Application of Coiled Coil Peptides in Liposomal Anticancer Drug Delivery Using a Zebrafish Xenograft Model," *ACS Nano*, vol. 10, no. 8, pp. 7428–7435, 2016.

[223] E. E. Oude Blenke, J. van den Dikkenberg, B. van Kolck, A. Kros, and E. Mastrobattista, "Coiled coil interactions for the targeting of liposomes for nucleic acid delivery," *Nanoscale*, vol. 8, no. 16, pp. 8955–8965, 2016.

[224] K. Wu, J. Yang, J. Liu, and J. Kopeček, "Coiled-coil based drug-free macromolecular therapeutics: In vivo efficacy," *J. Controlled Release*, vol. 157, no. 1, pp. 126–131, 2012.

[225] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity," *Science (80-. ).*, vol. 337, no. August, pp. 816–822, 2012.

[226] L. Cong, F. Ann Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang, "Multiplex Genome Engineering Using CRISPR/Cas Systems," *Science (80-. ).*, no. February, pp. 819–824, 2013.

[227] J. D. Sander and J. K. Joung, "CRISPR-Cas systems for editing, regulating and targeting genomes.," *Nat. Biotechnol.*, vol. 32, no. 4, pp. 347–55, 2014.

[228] X. Xiong, M. Chen, W. A. Lim, D. Zhao, and L. S. Qi, "CRISPR/Cas9 for Human Genome Engineering and Disease Research," *Annu. Rev. Genomics Hum. Genet.*, vol. 17, no. 1, pp. 131–154, 2016.

[229] N. Kolli, M. Lu, P. Maiti, J. Rossignol, and G. L. Dunbar, "Application of the gene editing tool, CRISPR-Cas9, for treating neurodegenerative diseases," *Neurochem. Int.*, vol. 112, pp. 187–196, 2018.

[230] E. R. Burnight, J. C. Giacalone, J. A. Cooke, J. R. Thompson, L. R. Bohrer, K. R. Chirco, A. V. Drack, J. H. Fingert, K. S. Worthington, L. A. Wiley, R. F. Mullins, E. M. Stone, and B. A. Tucker, "CRISPR-Cas9 genome engineering: Treating inherited retinal degeneration," *Prog. Retin. Eye Res.*, vol. 65, no. September 2017, pp. 28–49, 2018.

[231] Y. Lin, T. J. Cradick, M. T. Brown, H. Deshmukh, P. Ranjan, N. Sarode, B. M. Wile, P. M. Vertino, F. J. Stewart, and G. Bao, "CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences," *Nucleic Acids Res.*, vol. 42, no. 11, pp. 7473–7485, 2014.

[232] X. Liang, J. Potter, S. Kumar, Y. Zou, R. Quintanilla, M. Sridharan, J. Carte, W. Chen, N. Roark, S. Ranganathan, N. Ravinder, and J. D. Chesnut, "Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection," *J. Biotechnol.*, vol. 208, pp. 44–53, 2015.

[233] S. Ramakrishna, A. B. Kwaku Dad, J. Beloor, R. Gopalappa, S. K. Lee, and H. Kim, "Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA," *Genome Res.*, vol. 24, no. 6, pp. 1020–1027, 2014.

[234] J. G. Choi, Y. Dang, S. Abraham, H. Ma, J. Zhang, H. Guo, Y. Cai, J. G. Mikkelsen, H. Wu, P. Shankar, and N. Manjunath, "Lentivirus pre-packed with Cas9 protein for safer gene editing," *Gene Ther.*, vol. 23, no. 7, pp. 627–633, 2016.

[235] B. T. Staahl, M. Benekareddy, C. Coulon-Bainier, A. A. Banfal, S. N. Floor, J. K. Sabo, C. Urnes, G. A. Munares, A. Ghosh, and J. A. Doudna, "Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes," *Nat. Biotechnol.*, vol. 35, no. 5, pp. 431–434, 2017.

[236] K. Lee, M. Conboy, H. M. Park, F. Jiang, H. J. Kim, M. A. Dewitt, V. A. Mackley, K. Chang, A. Rao, C. Skinner, T. Shobha, M. Mehdipour, H. Liu, W. C. Huang, F. Lan, N. L. Bray, S. Li, J. E. Corn, K. Kataoka, J. A. Doudna, I. Conboy, and N. Murthy, "Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair," *Nat. Biomed. Eng.*, vol. 1, no. 11, pp. 889–901, 2017.

[237] J. A. Zuris, D. B. Thompson, Y. Shu, J. P. Guilinger, J. L. Bessen, J. H. Hu, M. L. Maeder, J. K. Joung, Z.-Y. Chen, and D. R. Liu, "Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo.," *Nat. Biotechnol.*, vol. 33, no. 1, pp. 73–80, 2014.

[238] C. M. Guenther, M. J. Brun, A. D. Bennett, M. L. Ho, W. Chen, B. Zhu, M. Lam, M. Yamagami, S. Kwon, N. Bhattacharya, D. Sousa, A. C. Evans, J. Voss, E. M. Sevick-Muraca, M. Agbandje-McKenna, and J. Suh, "Protease-activatable adeno-associated virus vector for gene delivery to damaged heart tissue," *Mol. Ther.*, vol. 0, no. 0, 2019.

[239] Y. Kitamoto, X. Yuan, Q. Y. Wu, D. W. Mccourt, and J. E. Sadler, "Enterokinase, the Initiator of Intestinal Digestion, Is a Mosaic Protease Composed of a Distinctive Assortment of Domains," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 16, pp. 7588–7592, 1994.

[240] H. M. Sassenfeld, "Engineering proteins for purification," *Trends Biotechnol.*, vol. 8, no. C, pp. 88–93, 1990.

[241] V. P. Reddy Chichili, V. Kumar, and J. Sivaraman, "Linkers in the structural biology of protein-protein interactions," *Protein Sci.*, vol. 22, no. 2, pp. 153–167, 2013.

[242] X. Chen, J. L. Zaro, and W. C. Shen, "Fusion protein linkers: Property, design and functionality," *Adv. Drug Deliv. Rev.*, vol. 65, no. 10, pp. 1357–1369, 2013.

[243] L. H. Vandenberghe, R. Xiao, M. Lock, J. Lin, M. Korn, and J. M. Wilson, "Efficient Serotype-Dependent Release of Functional Vector into the Culture Medium During Adeno-Associated Virus Manufacturing," *Hum. Gene Ther.*, vol. 21, no. 10, pp. 1251–1257, 2010.

[244] H. Mehdi, E. Ono, and K. C. Gupta, "Initiation of translation at CUG, GUG, and ACG codons in mammalian cells," *Gene*, vol. 91, pp. 173–178, 1990.

[245] M. Y. Chen, T. M. Robinson, and J. Suh, "Longer Inactivating Sequence in Peptide Lock Improves Performance of Synthetic Protease-Activatable Adeno-Associated Virus," *ACS Synth. Biol.*, vol. 8, p. acssynbio.8b00330, 2019.

[246] J. E. Rabinowitz, W. Xiao, and R. J. Samulski, "Insertional mutagenesis of AAV2 capsid and the production of recombinant virus.," *Virology*, vol. 265, no. 2, pp. 274–285, 1999.

[247] J. Judd, F. Wei, P. Q. Nguyen, L. J. Tartaglia, M. Agbandje-McKenna, J. J. Silberg, and J. Suh, "Random Insertion of mCherry Into VP3 Domain of Adeno-associated Virus Yields Fluorescent Capsids With no Loss of Infectivity.," *Mol. Ther. Nucleic Acids*, vol. 1, p. e54, jan 2012.

[248] J. F. Wright, "Manufacturing and characterizing AAV-based vectors for use in clinical studies," *Gene Ther.*, vol. 15, no. 11, pp. 840–848, 2008.

[249] D. Reyon, S. Q. Tsai, C. Khgayter, J. A. Foden, J. D. Sander, and J. K. Joung, "FLASH assembly of TALENs for high-throughput genome editing," *Nat. Biotechnol.*, vol. 30, no. 5, pp. 460–465, 2012.

[250] D. Y. Guschin, A. J. Waite, G. E. Katibah, J. C. Miller, M. C. Holmes, and E. J. Rebar, "A Rapid and General Assay for Monitoring Endogenous Gene Modification," *Methods Mol. Biol.*, vol. 649, pp. 247–256, 2010.

[251] S. Russell, J. Bennett, J. A. Wellman, D. C. Chung, Z. F. Yu, A. Tillman, J. Wittes, J. Pappas, O. Elci, S. McCague, D. Cross, K. A. Marshall, J. Walshire, T. L. Kehoe, H. Reichert, M. Davis, L. Raffini, L. A. George, F. P. Hudson, L. Dingfield, X. Zhu, J. A. Haller, E. H. Sohn, V. B. Mahajan, W. Pfeifer, M. Weckmann, C. Johnson, D. Gewaily, A. Drack, E. Stone, K. Wachtel, F. Simonelli, B. P. Leroy, J. F. Wright, K. A. High, and A. M. Maguire, "Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial," *Lancet*, vol. 390, no. 10097, pp. 849–860, 2017.

[252] D. Grimm, K. Pandey, H. Nakai, T. A. Storm, and M. A. Kay, "Liver Transduction with Recombinant Adeno-Associated Virus Is Primarily Restricted by Capsid Serotype Not Vector Genotype," *J. Virol.*, vol. 80, no. 1, pp. 426–439, 2006.

[253] G. Gao, Y. Lu, R. Calcedo, R. L. Grant, P. Bell, L. Wang, J. Figueredo, M. Lock, and J. M. Wilson, "Biology of AAV serotype vectors in liver-directed gene transfer to nonhuman primates," *Mol. Ther.*, vol. 13, no. 1, pp. 77–87, 2006.

[254] H. Nakai, S. R. Yant, T. A. Storm, S. Fuess, L. Meuse, and M. A. Kay, "Extrachromosomal Recombinant Adeno-Associated Virus Vector Genomes Are Primarily Responsible for Stable Liver Transduction In Vivo," *J. Virol.*, vol. 75, no. 15, pp. 6969–6976, 2001.

[255] D. Duan, P. Sharma, J. Yang, Y. Yue, L. Dudus, Y. Zhang, K. J. Fisher, and J. F. Engelhardt, "Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue.," *J. Virol.*, vol. 72, no. 11, pp. 8568–77, 1998.

[256] W. C. Manning, S. Zhou, M. P. Bland, J. A. Escobedo, and V. Dwarki, "Transient Immunosuppression Allows Transgene Expression Following Readministration of Viral Vectors," *Hum. Gene Ther.*, vol. 9, pp. 477–485, 1998.

[257] C. L. Halbert, T. A. Standaert, C. B. Wilson, and A. D. Miller, "Successful readministration of adeno-associated virus vectors to the mouse lung requires transient immunosuppression during the initial exposure.," *J. Virol.*, vol. 72, no. 12, pp. 9795–805, 1998.

[258] C. L. Halbert, E. A. Rutledge, J. M. Allen, D. W. Russell, and A. D. Miller, "Repeat transduction in the mouse lung by using adeno-associated virus vectors with different serotypes.," *J. Virol.*, vol. 74, no. 3, pp. 1524–1532, 2000.

[259] C. Hinderer, N. Katz, E. L. Buza, C. Dyer, T. Goode, P. Bell, L. K. Richman, and J. M. Wilson, "Severe Toxicity in Nonhuman Primates and Piglets Following High-Dose Intravenous Administration of an Adeno-Associated Virus Vector Expressing Human SMN," *Hum. Gene Ther.*, vol. 29, no. 3, pp. 285–298, 2018.

[260] M. L. Ho, B. a. Adler, M. L. Torre, J. J. Silberg, and J. Suh, "SCHEMA computational design of virus capsid chimeras: calibrating how genome packaging, protection, and transduction correlate with calculated structural disruption.," *ACS Synth. Biol.*, vol. 2, pp. 724–33, dec 2013.

[261] D. S. Ojala, S. Sun, J. L. Santiago-Ortiz, M. G. Shapiro, P. A. Romero, and D. V. Schaffer, "In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the Subventricular Zone," *Mol. Ther.*, 2017.

[262] K. Adachi, T. Enoki, Y. Kawano, M. Veraz, and H. Nakai, "Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing.," *Nat. Commun.*, vol. 5, no. June 2011, p. 3075, 2014.

[263] E. Zinn, S. Pacouret, V. Khaychuk, H. T. Turunen, L. S. Carvalho, E. Andres-Mateos, S. Shah, R. Shelke, A. C. Maurer, E. Plovie, R. Xiao, and L. H. Vandenberghe, "In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector," *Cell Rep.*, pp. 1–13, 2015.

[264] L. D. Landegger, B. Pan, C. Askew, S. J. Wassmer, S. D. Gluck, A. Galvin, R. Taylor, A. Forge, K. M. Stankovic, J. R. Holt, and L. H. Vandenberghe, "A synthetic AAV vector enables safe and efficient gene transfer to the mammalian inner ear," *Nat. Biotechnol.*, no. February 2016, 2017.

[265] M. G. Mateu, *Structure and Physics of Viruses*, vol. 68. Springer, 2013.

[266] A. Zlotnick, "To build a virus capsid: An equilibrium model of the self assembly of polyhedral protein complexes," *J. Mol. Biol.*, no. 241, pp. 59–67, 1994.

[267] S. Katen and A. Zlotnick, "The Thermodynamics of Virus Capsid Assembly," *Methods Enzymol.*, vol. 6879, no. 08, pp. 1–19, 2009.

[268] N. Onuchic and P. G. Wolynes, "Theory of protein folding," *Curr. Opin. Struct. Biol.*, pp. 70–75, 2004.

[269] J. D. Bryngelson and P. G. Wolynes, "Spin glasses and the statistical mechanics of protein folding," *Proc. Natl. Acad. Sci.*, vol. 84, no. November, pp. 7524–7528, 1987.

[270] D. U. Ferreiro, J. A. Hegler, E. A. Komives, and P. G. Wolynes, "On the role of frustration in the energy landscapes of allosteric proteins," *Proc. Natl. Acad. Sci.*, vol. 108, no. 9, pp. 3499–3503, 2011.

[271] P. I. Zhuravlev and G. A. Papoian, "Protein functional landscapes, dynamics , allostery: a tortuous path towards a universal theoretical framework," *Q. Rev. Biophys.*, vol. 3, pp. 295–332, 2010.

[272] R. G. Parra, N. P. Schafer, L. G. Radusky, M. Y. Tsai, A. B. Guzovsky, P. G. Wolynes, and D. U. Ferreiro, "Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W356–W360, 2016.

[273] J. Onuchic, Z. Schulten, and P. Wolynes, "THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective," *Annu. Rev. Phys. Chem.*, vol. 48, no. 1, pp. 545–600, 1997.

[274] W. Zheng, N. P. Schafer, A. Davtyan, G. A. Papoian, and P. G. Wolynes, "Predictive energy landscapes for protein-protein association," *Proc. Natl. Acad. Sci.*, vol. 109, no. 47, pp. 19244–19249, 2012.

[275] D. Potoyan, W. Zheng, E. Komives, and P. Wolynes, "Molecular stripping in the NF$\kappa$B/I$\kappa$B/DNA genetic regulatory network," *bioRxiv*, vol. 113, no. 1, p. 027912, 2015.

[276] A. Arkhipov, P. L. Freddolino, and K. Schulten, "Stability and Dynamics of Virus Capsids Described by Coarse-Grained Modeling," *Structure*, vol. 14, no. 12, pp. 1767–1777, 2006.

[277] J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu, and K. Schulten, "Molecular dynamics simulations of large macromolecular complexes," *Curr. Opin. Struct. Biol.*, vol. 31, pp. 64–74, 2015.

[278] M. Sitharam and M. Agbandje-McKenna, "Modeling Virus Self-Assembly Pathways," *J. Comput. Biol.*, vol. 13, no. 6, pp. 1232–12651. Sitharam M, Agbandje–McKenna M (2006), 2006.

[279] T. Reddy and M. S. Sansom, "Computational virology: from the inside out," *Biochim. Biophys. Acta - Biomembr.*, 2016.

[280] M. F. Hagan and R. Zandi, "Recent advances in coarse-grained modeling of virus assembly," *Curr. Opin. Virol.*, vol. 18, pp. 36–43, 2016.

[281] A. Callaway, D. Giesman-Cookmeyer, E. T. Gillock, T. L. Sit, and S. A. Lommel, "The multifunctional capsid proteins of plant RNA viruses," *Annu. Rev. Phytopathol.*, 2001.

[282] M. Tu, F. Liu, S. Chen, M. Wang, and A. Cheng, "Role of capsid proteins in parvoviruses infection," *Virol. J.*, 2015.

[283] I. N. Shindyalov, N. A. Kolchanov, and C. Sander, "Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations?," *Protein Eng.*, vol. 7, no. 3, pp. 349–358, 1994.

[284] U. Gobel, C. Sander, R. Schneider, and A. Valencia, "Correlated Mutations and Residue Contacts in Proteins," *Proteins Struct. Funct. Genet.*, vol. 317, pp. 309–317, 1994.

[285] G. M. Süel, S. W. Lockless, M. a. Wall, and R. Ranganathan, "Evolutionarily conserved networks of residues mediate allosteric communication in proteins.," *Nat. Struct. Biol.*, vol. 10, no. 1, pp. 59–69, 2003.

[286] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, "Identification of direct residue contacts in protein-protein interaction by message passing.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 1, pp. 67–72, 2009.

[287] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, and C. Sander, "Direct-coupling analysis of residue coevolution captures native contacts across many protein families," *PNAS*, 2011.

[288] D. S. Marks, T. A. Hopf, and C. Sander, "Protein structure prediction from sequence variation," *Nat. Publ. Gr.*, vol. 30, no. 11, pp. 1072–1080, 2012.

[289] F. Morcos, N. P. Schafer, R. R. Cheng, J. N. Onuchic, and P. G. Wolynes, "Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection," *Proc. Natl. Acad. Sci.*, vol. 111, no. 34, pp. 12408–12413, 2014.

[290] F. Morcos, B. Jana, T. Hwa, and J. N. Onuchic, "Coevolutionary signals across protein lineages help capture multiple protein conformations," *Proc. Natl. Acad. Sci.*, vol. 110, no. 51, pp. 20533–20538, 2013.

[291] J. S. Boyd, R. R. Cheng, M. L. Paddock, C. Sancar, F. Morcos, and S. Golden, "A Combined Computational and Genetic Approach Uncovers Network Interactions of the Cyanobacterial Circadian Clock," *J. Bacteriol.*, vol. 198, no. 18, pp. 2439–2447, 2016.

[292] A. R. R. Cheng, O. Nordesjö, R. L. Hayes, H. Levine, and S. C. Flores, "Connecting the sequence - space of bacterial signaling proteins to phenotypes using coevolutionary landscapes," *Mol. Biol. Evol.*, pp. 1–36, 2016.

[293] R. R. Cheng, E. Haglund, N. S. Tiee, F. Morcos, H. Levine, J. A. Adams, P. A. Jennings, and J. N. Onuchic, "Designing bacterial signaling interactions with coevolutionary landscapes," *PLoS One*, vol. 13, no. 8, pp. 1–16, 2018.

[294] R. N. Dos Santos, F. Morcos, B. Jana, A. D. Andricopulo, and J. N. Onuchic, "Dimeric interactions and complex formation using direct coevolutionary couplings.," *Sci. Rep.*, vol. 5, p. 13652, 2015.

[295] G. Uguzzoni, S. John, F. Oteri, A. Schug, H. Szurmant, and M. Weigt, "Large-scale identification of coevolution signals across homo-oligomeric protein interfaces by direct coupling analysis," *Proc. Natl. Acad. Sci.*, 2017.

[296] D. Altschuht, A. M. Leskx, A. C. Bloomer, and A. Klug, "Correlation of Co-ordinated Amino Acid Substitutions with Function in Viruses Related to Tobacco Mosaic Virus," *J. Mol. Biol.*, pp. 693–707, 1987.

[297] G. Li, K. Theys, J. Verheyen, R. Khouri, S. Piampongsant, M. Eusébio, J. Ramon, and A.-m. Vandamme, "A new ensemble coevolution system for detecting HIV-1 protein coevolution," *Biol. Direct*, pp. 1–20, 2015.

[298] R. Champeimont, E. Laine, S.-w. Hu, and F. Penin, "Coevolution analysis of Hepatitis C virus genome to identify the structural and functional dependency network of viral proteins," *Sci. Rep.*, no. October 2015, pp. 1–20, 2016.

[299] D. U. Ferreiro, J. A. Hegler, E. A. Komives, and P. G. Wolynes, "Localizing frustration in native proteins and protein assemblies.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 50, pp. 19819–24, 2007.

[300] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, 2000.

[301] J. Mintseris, K. Wiehe, B. Pierce, R. Anderson, R. Chen, J. Janin, and Z. Weng, "Protein-Protein Docking Benchmark 2.0: An update," *Proteins Struct. Funct. Genet.*, vol. 60, no. 2, pp. 214–216, 2005.

[302] J. K. Noel, F. Morcos, and J. N. Onuchic, "Sequence co-evolutionary information is a natural partner to minimally-frustrated models of biomolecular dynamics," *F1000Research*, vol. 5, no. 0, pp. 1–7, 2016.

[303] Y. Yang, J. Gao, J. Wang, R. Heffernan, J. Hanson, K. Paliwal, and Y. Zhou, "Sixty-five years of the long march in protein secondary structure prediction: The final stretch?," *Brief. Bioinform.*, vol. 19, no. 3, pp. 482–494, 2018.

[304] G. A. Papoian and P. G. Wolynes, "AWSEM-MD," in *Coarse-Grained Model. Biomol.* (G. A. Papoian, ed.), ch. 4, pp. 121–190, CRC Press, 1st editio ed., 2017.

[305] M. F. Hagan and D. Chandler, "Dynamic pathways for viral capsid assembly," *Biophys. J.*, vol. 91, no. 1, pp. 42–54, 2006.

[306] M. F. Hagan, "Controlling viral capsid assembly with templating," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 77, no. 5, pp. 1–11, 2008.

[307] Q. Xie, W. Bu, S. Bhatia, J. Hare, T. Somasundaram, A. Azzi, and M. S. Chapman, "The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 16, pp. 10405–10410, 2002.

[308] S. F. Cotmore, M. Agbandje-McKenna, J. A. Chiorini, D. V. Mukha, D. J. Pintel, J. Qiu, M. Soderlund-Venermo, P. Tattersall, P. Tijssen, D. Gatherer, and A. J. Davison, "The family Parvoviridae," *Arch. Virol.*, vol. 159, no. 5, pp. 1239–1247, 2014.

[309] B. R. Schultz and J. S. Chamberlain, "Recombinant adeno-associated virus transduction and integration.," *Mol. Ther.*, vol. 16, no. 7, pp. 1189–1199, 2008.

[310] B. Jana, F. Morcos, and J. N. Onuchic, "From structure to function: the convergence of structure based models and co-evolutionary information.," *Phys. Chem. Chem. Phys.*, vol. 16, no. October 2016, pp. 6496–6507, 2014.

[311] S. Tamir, S. Rotem-Bamberger, C. Katz, F. Morcos, K. L. Hailey, J. A. Zuris, C. Wang, A. R. Conlan, C. H. Lipper, M. L. Paddock, R. Mittler, J. N. Onuchic, P. A. Jennings, A. Friedler, and R. Nechushtai, "Integrated strategy reveals the protein interface between cancer targets Bcl-2 and NAF-1," *Proc. Natl. Acad. Sci.*, vol. 111, no. 14, pp. 5177–5182, 2014.

[312] R. R. Cheng, F. Morcos, H. Levine, and J. N. Onuchic, "Toward rationally redesigning bacterial two-component signaling systems using coevolutionary information," *Proc. Natl. Acad. Sci.*, vol. 111, no. 5, pp. E563–E571, 2014.

[313] L. Govindasamy, E. Padron, R. McKenna, N. Muzyczka, N. Kaludov, J. A. Chiorini, and M. Agbandje-McKenna, "Structurally mapping the diverse phenotype of adeno-associated virus serotype 4.," *J. Virol.*, vol. 80, no. 23, pp. 11556–70, 2006.

[314] L. Govindasamy, M. A. Dimattia, L. Brittney, S. Halder, R. Mckenna, J. A. Chiorini, N. Muzyczka, and S. Zolotukhin, "Structural Insights into Adeno-Associated Virus Serotype 5," *J. Virol.*, vol. 87, no. 20, pp. 11187–11199, 2013.

[315] R. Ng, L. Govindasamy, B. L. Gurda, R. McKenna, O. G. Kozyreva, R. J. Samulski, K. N. Parent, T. S. Baker, and M. Agbandje-McKenna, "Structural characterization of the dual glycan binding adeno-associated virus serotype 6.," *J. Virol.*, vol. 84, no. 24, pp. 12945–12957, 2010.

[316] H.-J. Nam, M. D. Lane, E. Padron, B. Gurda, R. McKenna, E. Kohlbrenner, G. Aslanidi, B. Byrne, N. Muzyczka, S. Zolotukhin, and M. Agbandje-McKenna, "Structure of adeno-associated virus serotype 8, a gene therapy vector.," *J. Virol.*, vol. 81, no. 22, pp. 12260–12271, 2007.

[317] M. A. DiMattia, H. J. Nam, K. Van Vliet, M. Mitchell, A. Bennett, B. L. Gurda, R. McKenna, N. H. Olson, R. S. Sinkovits, M. Potter, B. J. Byrne, G. Aslanidi, S. Zolotukhin, N. Muzyczka, T. S. Baker, and M. Agbandje-McKenna, "Structural insight into the unique properties of adeno-associated virus serotype 9," *J Virol*, vol. 86, no. 12, pp. 6947–6958, 2012.

[318] A. Davtyan, N. P. Schafer, W. Zheng, C. Clementi, P. G. Wolynes, and G. A. Papoian, "AWSEM-MD: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing," *J. Phys. Chem. B*, vol. 116, no. 29, pp. 8494–8503, 2012.

[319] X. Xiao, J. Li, and R. J. Samulski, "Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus.," *J. Virol.*, vol. 72, no. 3, pp. 2224–32, 1998.

[320] I. W. Hamley, "Small Bioactive Peptides for Biomaterials Design and Therapeutics," *Chem. Rev.*, vol. 117, pp. 14015–14041, 2017.

[321] B. Yameen, W. I. Choi, C. Vilos, A. Swami, J. Shi, and O. C. Farokhzad, "Insight into nanoparticle cellular uptake and intracellular targeting," *J. Controlled Release*, vol. 190, pp. 485–499, 2014.

[322] M. Castellanos, R. Pérez, A. Rodríguez-Huete, E. Grueso, J. M. M. Almendral, and M. G. G. Mateu, "A slender tract of glycine residues is required for translocation of the VP2 protein N-terminal domain through the parvovirus MVM capsid channel to initiate infection.," *Biochem. J.*, vol. 455, no. 1, pp. 87–94, 2013.

[323] P. Wang, L. Zhang, Y. Xie, N. Wang, R. Tang, W. Zheng, and X. Jiang, "Genome Editing for Cancer Therapy: Delivery of Cas9 Protein/sgRNA Plasmid via a Gold Nanocluster/Lipid Core–Shell Nanocarrier," *Adv. Sci.*, vol. 4, no. 11, 2017.