TOWARDS A STRUCTURAL UNDERSTANDING OF ADENO-ASSOCIATED VIRUS SEROTYPE 2 AND ITS RECOGNITION BY ANTIBODIES

By

Dustin McCraw

A Dissertation

Presented to the Department of Biochemistry and Molecular Biology
And the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

June 2012

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the PhD dissertation of

Dustin McCraw

has been approved

_____
Mentor/Advisor

_____
Member

_____
Member

_____
Member

_____
Member

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS

5% HMG	100 mM HEPES, 50 mM magnesium chloride, 5% glycerol
AAV	adeno-associated virus
BS3	bis[sulfosuccinimidyl] suberate
BSA	bovine serum albumin
Cap	capsid
CCS	Cosmic Calf Serum
CPV	canine parvovirus
CTF	contrast transfer function
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
EM	electron microscopy
Fab	fragment (antigen binding)
FCS	fetal calf serum
FPV	feline panleukopenia virus
HSPG	heparan sulfate proteoglycan
ITR	inverted terminal repeat
JMEM	Joklik's Modification of Eagle's Minimal Essential Medium
kb	kilobase pair(s)
LB media	lysogeny broth
MAb	monoclonal antibody
MVM	Minute virus of mice

NCS	non-crystallographic symmetry
NPC	nuclear pore complex
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PLA2	phospholipase A2
rAAV	recombinant AAV
Rep	replication
RCF	relative centrifugal force
RMS(D)	root mean square (deviation)
RPMI	Roswell Park Memorial Institute
SOC Media	super-optimal broth with catabolite repression
ssDNA	single-stranded DNA
TBS-T	tris-buffered saline and Tween-20
ssDNA	single-stranded deoxyribonucleic acid
vdW	van der Waals
VP	viral protein
VP1u	viral protein 1 unique
VR	variable region

ABSTRACT

Adeno-associated virus (AAV) is a leading candidate as a gene therapy vector, but as with other vectors, one of the limitations is a host neutralizing immune response. The most extensively characterized immune interaction is that between AAV serotype 2 and neutralizing monoclonal antibody (MAb) A20.

The monovalent A20 Fab fragment and AAV-2 were initially isolated for structural studies. A20 monoclonal antibody is confirmed to have high binding affinity to AAV-2, however A20 Fab is shown to have only low binding affinity and a complex could not be isolated. Alternative production methods show that A20 Fab' exhibits a relatively high binding affinity to AAV-2.

A *cryo*-electron microscopy (EM) structure of AAV-2 complexed with the Fab' fragment of A20 has been determined to 8.5 Å resolution. The binding footprint is determined through fitting the *cryo*-EM reconstruction with a homology model following sequencing of the variable domain, and provides a structural basis for integrating diverse prior epitope mappings. Direct visualization sheds new light on prior experiments, resolving inconsistencies in proposed components of the epitope. The A20 footprint on AAV-2 extends from the plateau, a region of moderate sequence diversity implicated in some of the prior studies, to the side of the spike, and into the conserved canyon, covering a much larger area than anticipated. Comparison with structures of binding and non-binding serotypes indicates that recognition depends on a combination of subtle serotype-specific features.

Separation of the neutralizing epitope from the heparan sulfate cell attachment site encourages attempts to develop immune-resistant vectors that can still bind to target cells.

Additional assessment of the electrostatic surface and the free energy of solvation of the binding site show that there is no singular distinguishing characteristic between AAV serotypes that bind to A20 and those which do not. The current structure suggests that A20 recognition is therefore moderated by several relatively small modulations in serotype differences at the binding site.

INTRODUCTION

PARVOVIRUSES

Parvoviruses are among some the smallest known viruses and are widespread in nature. Adeno-associated virus was originally discovered in 1965 and is not associated with any disease (Atchison, Casto, and Hammon, 1965). Parvovirus B19 was discovered in 1975 and is pathogenic to humans (Cossart et al., 1975). They may be replication deficient, such as Adeno-associated virus, and need the coinfection of a helper virus (Agbandje-McKenna and Kleinschmidt, 2011), or they may be replication autonomous such as parvovirus B19 (Shneerson, Mortimer, and Vandervelde, 1980).

Family Parvoviridae

The family *Parvoviridae* consists of small viruses of about 260-300 Å in diameter (Kerr, 2006). The viruses exhibit icosahedral, T=1 symmetry, where 60 identical subunits each occupy a quasi-equivalent position (Kerr, 2006). They are non-enveloped, containing only a single protein shell, and contain an approximately ~5 kb genome of ssDNA (Kerr, 2006). The 60 subunits of the capsid protein contain usually 8 strands of an antiparallel, β-barrel motif (Kerr, 2006).

The family *Parvoviridae* contains the subfamilies *Densoviridae* and *Parvovirinae*. Viruses of the subfamily *Densoviridae* infect invertebrate hosts. *Densoviridae* is further divided into the genera *Brevidensovirus*, *Densovirus*, *Iteravirus*, and

Pefudensovirus. Subfamily *Parvovirinae* consists of the genera *Amdovirus*, *Bocavirus*, *Dependovirus*, *Erythrovirus*, and *Parvovirus*. Viruses of the genus *Parvovirus* are autonomous viruses and are dependent on the cell undergoing S-phase for replication (Rhode, 1973; Tattersall, 1972). Additionally, parvoviruses possess relatively narrow host and tissue specificity and, due to their autonomous nature, are significant pathogens (Berns and Labow, 1987). Viruses of the genus *Depenovirus,* however, are not autonomous. While not dependent on S-phase for replication, dependovirus replication is dependent on the helper functions of a co-infecting virus such as Adenovirus, Herpesevirus, Papillomavirus, or Vaccinia virus (King, 2012). Relative to parvoviruses, dependoviruses may possess significantly wider host range (Berns and Labow, 1987).

ADENO-ASSOCIATED VIRUS

Adeno-associated virus is a species of the genus *Dependovirus* (King, 2012). It is a leading candidate as a gene therapy vector in part due to its nonpathogenic nature and constant integration site (Agbandje-McKenna and Kleinschmidt, 2011). AAV is a non-enveloped virus that contains a single protein capsid, which encases a single stranded DNA genome (Caspar and Klug, 1962; Luo et al., 1988). The capsid is composed of 60 protein subunits related by icosahedral symmetry (Xie et al., 2002).

Genome

The AAV genome has inverted terminal repeats (ITRs) at both 3' and 5' ends (Muzyczka and Berns, 2001) which function in DNA replication, regulation of gene expression, site specific integration, rescue, and packaging (Qing et al., 1997; Wang and Srivastava, 1997). There are two open reading frames (ORFs) that together encode 4 overlapping replication proteins and 3 overlapping capsid proteins (Rutledge, Halbert, and Russell, 1998). A diagram of the AAV-2 genome can be seen in Figure 1.

Expression of Rep 78 and Rep 68 is controlled by the P5 promoter and expression of Rep 52 and Rep 40 is controlled by the P19 promoter. A splice site common to both promoters creates two *rep* proteins via alternative splicing for each mRNA. The four *rep* proteins have several functions: transcriptional regulation, helicase activity, endonuclease activity, and site-specific integration (*vide infra*) (Muzyczka and Berns, 2001).

The *cap* open reading frame encodes the three capsid proteins. Expression of the capsid proteins is controlled by the P40 promoter. The three capsid proteins; VP-1, VP-2, and VP-3 are present in a respective ratio of 1:1:10, and have molecular weights of 87 kD, 73 kD, and 62 kD, differing (only) in N-terminal extensions of 137 residues (VP-1) and 64 residues (VP-2) relative to VP-3 (Hoque et al., 1999). Alternative splicing creates both the mRNA encoding the VP-1 transcript and the mRNA encoding the VP-2 and VP-3 transcript. VP-2 has an atypical ACG start codon and VP-3 has a more classical AUG start codon. VP-3 proteins are produced at a

higher ratio than VP-2 due to its initiation at an AUG codon. The ratio of VP-1 proteins to VP-2/-3 is determined by the level of alternative splicing, which is in turn determined by the availability of helper functions (*vide infra*). It is not clear whether VP-1 & 2 substitute for VP-3 at specific locations in the otherwise 60-fold symmetric capsid. The N-terminal domain for VP-1 has phospholipase activity (Girod et al., 2002), while VP-2 appears to function in nuclear localization (Hoque et al., 1999; Wu et al., 2000).

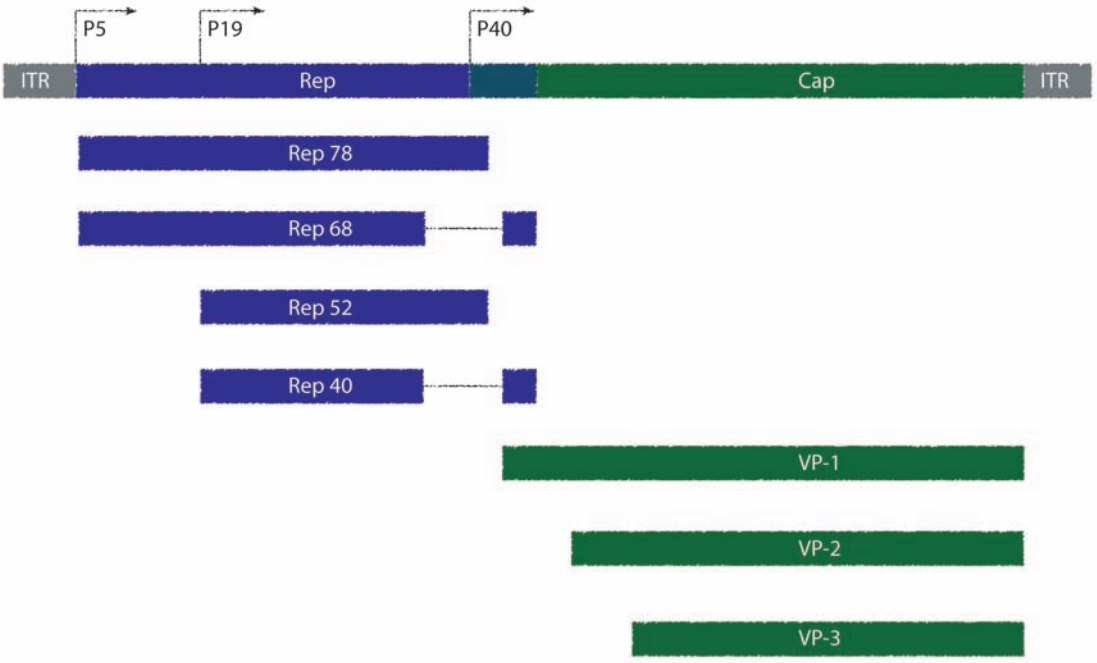


Figure 1. The AAV-2 genome consists of two open reading frames enclosed by two ITR sequences at the termini. The *rep* open reading frame has two promoters (P5 and P19) and uses alternative splicing to encodes four *rep* proteins. The *cap* open reading frame has the P40 promoter and encodes three viral capsid proteins through the use of alternative splicing and alternative initiation. Figure adapted from (Mouw and Pintel, 2000).

Uptake and Endosomal Trafficking

Initial AAV infection is latent. Replication requires co-infection by a helper-virus, usually adenovirus (Rivadeneira, 1998). AAV-2 and AAV-3 use heparan sulfate proteoglycans (HSPGs) as a primary receptor for cell entry (Summerford and Samulski, 1998; Walters et al., 2001). In contrast, AAV-1, AAV-4, AAV-5, and AAV-6 require sialic acid binding for efficient transduction, although AAV-6 can also bind to heparin (Kaludov et al., 2001; Wu et al., 2006). It is generally believed that AAV then interacts with a secondary receptor. Several secondary receptors have been reported to be involved with internalization of AAV including integrin $\alpha V\beta 5$, integrin $\alpha 5\beta 1$, fibroblast growth factor receptor 1, hepatocyte growth factor receptor, and 37/67 kDa laminin receptor (Agbandje-McKenna and Kleinschmidt, 2011). After interaction with a secondary receptor, AAV is engulfed via endocytosis where it is internalized and transported into the cell via endosomes (Summerford and Samulski, 1998). It is debated whether or not AAV passes through early endosomes (Bartlett, Wilcher, and Samulski, 2000; Xiao et al., 2002) or late endosomes (Douar et al., 2001; Hansen, Qing, and Srivastava, 2001) before it exits. Phospholipase A2 (PLA2) is reported to be necessary to breach the endosome for MVM (Farr, Zhang, and Tattersall, 2005). The PLA2 enzyme has been shown to be activated by a pH of approximately 5.5 (Ito et al., 2002). A domain with homology to phospholipase A2 (PLA2) has been reported to play a crucial role in the exit of AAV from the endosome (Farr, Zhang, and Tattersall, 2005; Stahnke et al., 2011; Suikkanen et al., 2003). The PLA2 domain is present in the unique N-terminal

region, VP1u of the VP1 subunit and is typically not exposed on the surface of AAV (Zadori et al., 2001). VP1u may be activated via a drop in the pH of the endosome, where it is exposed through the 5-fold axis of symmetry, and facilitate AAV-2 exit from the endosome (Bartlett, Wilcher, and Samulski, 2000; Bleker, Sonntag, and Kleinschmidt, 2005; Douar et al., 2001).

AAV has been shown to interact with microtubule-associated proteins (Kelkar et al., 2006; Zhao et al., 2006). Several reports suggest that the trafficking of AAV to the nucleus is mediated by microtubules (Hirosue et al., 2007; Sanlioglu et al., 2000; Seisenberger et al., 2001). It has been reported that AAV accumulated perinuclearly within 30 minutes of the onset of endocytosis (Samulski, Sally, and Muzyczka, 1999). AAV can be detected within the nucleus within 2 hours of infection, suggesting that AAV uncoats after having passed through the nuclear membrane (Samulski, 2000).

Precisely how AAV-2 enters the nucleus is not currently known and is an active area of research. AAV-2 is small enough to enter through the nuclear pore complex (NPC) and it has been suggested that entry into the cell nucleus occurs through an NPC dependent pathway (Bartlett, Wilcher, and Samulski, 2000). However, it has also been reported that the autonomous parvovirus MVM creates fissures in the nuclear membrane allowing viral entry through a mechanism independent of the NPC (Cohen and Pante, 2005). It has been suggested that the PLA2 domain may be instrumental in creating the fissures for nuclear entry (Cohen et al., 2006). AAV-2 may use a similar mechanism, as the PLA2 domain is also present in VP1 (Agbandje-

McKenna and Kleinschmidt, 2011). However, this remains controversial as the relatively high MOIs used to achieve membrane fissures would disrupt ionic potentials within the cell. It is not known whether, at more moderate MOI, disruption of the nuclear membrane could be more limited while still useful to AAV. Finally, AAV-2 capsid uncoating likely occurs in the nucleus (Bartlett, Wilcher, and Samulski, 2000; Sanlioglu et al., 2000; Xiao et al., 2002).

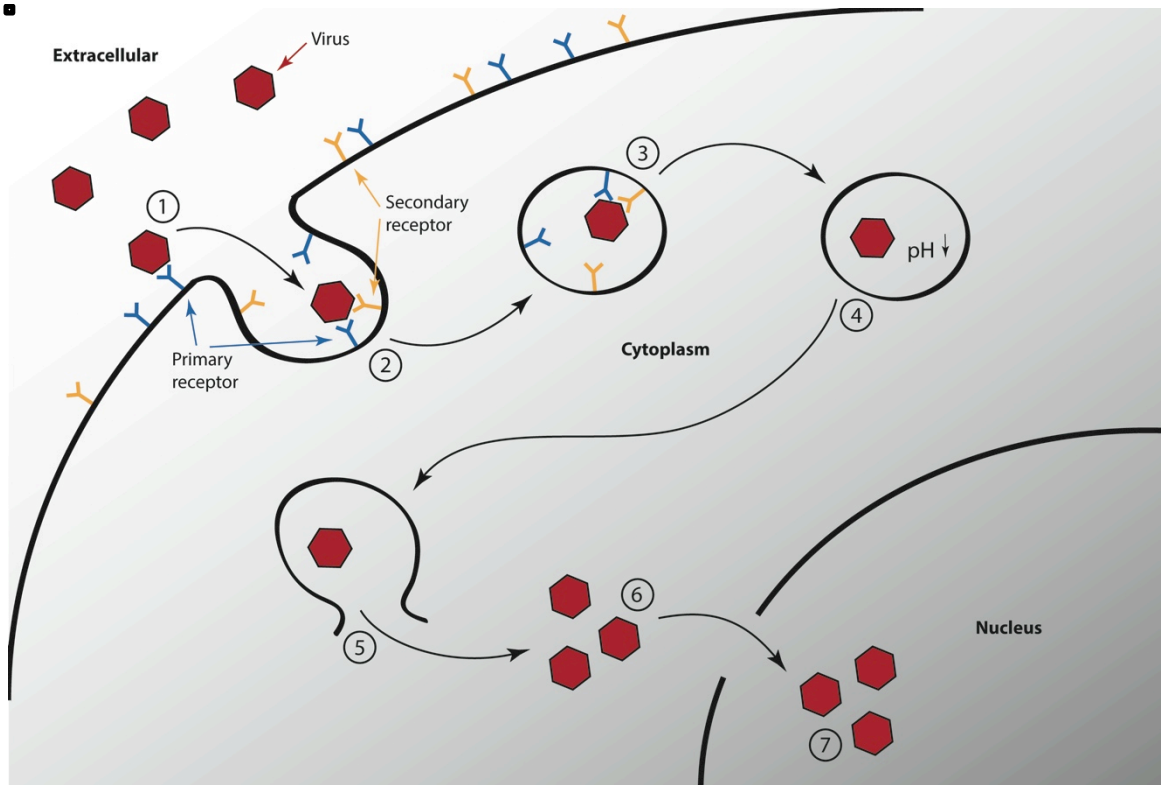


Figure 2. The uptake and endosomal trafficking pathway of AAV-2. (1) AAV-2 initially interacts with the primary receptor HSPG prior to cell entry. (2) AAV-2 interacts with an additional secondary receptor. (3) AAV-2 is engulfed and taken into the cell via endocytosis. (4) The pH of the endosome drops and (5) AAV-2 escapes from the endosome. (6) AAV-2 enters the cell cytoplasm before (7) entering the cell nucleus where uncoating occurs. Figure adapted from (Mitchell, 2008).

DNA Replication, Transcription, and Translation

ssDNA from AAV-2 is released from the capsid after uncoating, where it must be converted into dsDNA (Ferrari FK, 1996; Fisher KJ, 1996; Thomas et al., 2004). Second-strand synthesis is required for episomal expression and for integration of the dsDNA into the host's genome (Weitzman and Linden, 2011). Specifically, AAV undergoes viral second-strand DNA synthesis for both the positive and negative sense strand (Zhong et al., 2008). Both host proteins and *rep* proteins are necessary for replication of the complimentary strand. The second strand is created via a rolling hairpin mechanism. The dsDNA then integrates at a specific site on chromosome 19 where it remains until helper functions are available for replication (Kotin et al., 1990; Samulski et al., 1991).

When helper functions are not available, the AAV-2 genome only expresses limited amounts of Rep 68 and Rep 78 (Pereira, McCarty, and Muzyczka, 1997). Additionally, all four *rep* proteins have been shown to inhibit transcription of more *rep* proteins whenever helper functions are not present (Pereira, McCarty, and Muzyczka, 1997). When helper functions are available from a co-infecting adenovirus, the P5, P19, and P40 promoters are activated, and all four *rep* proteins and all three *cap* proteins are subsequently expressed (Pereira, McCarty, and Muzyczka, 1997).

AAV-2 DNA replication is accomplished through a rolling hairpin mechanism. Polymerase δ is part of the host replication machinery and is believed to be involved in AAV-2 DNA replication (Nash, Chen, and Muzyczka, 2008). Helper functions may

also replace some functionality of the host replication machinery during replication of the AAV-2 genome (Slanina et al., 2006). The ITRs reside on both sides of the *rep* and *cap* genes. The ITRs are self-annealing and contain a 3'-OH that facilitates DNA replication (Weitzman and Linden, 2011). However, the remaining ITR must still be copied. To accomplish this, *rep* proteins bind to the Rep binding site within the ITRs (Slanina et al., 2006). The 3' end of the AAV template is then copied, with replication terminating at the terminal resolution site. Replication termination leaves a nick in the dsDNA at a 3'-OH, where Rep helicase is required to leave the 3'-OH single-stranded and accessible for replication (Owens et al., 1993).

Splicing of *rep* and *cap* proteins is responsible for the production of five (Rep 68, Rep 40, VP-1, VP-2, and VP-3) of the seven AAV-2 encoded proteins. The ratio of Rep 78 to Rep68 and the ratio of Rep 52 to Rep 40 is dependent on the level of alternative splicing (Mouw and Pintel, 2000). Similarly, the ratio of VP-1 to VP-2 and VP-3 is also dependent on the level of alternative splicing (Mouw and Pintel, 2000). The level of splicing is determined by the availability of helper functions (Mouw and Pintel, 2000). Without helper-virus co-infection, splice site acceptors A1 and A2 are used equally (Mouw and Pintel, 2000). However, with helper-virus co-infection, splice site acceptor A2 to A1 utilization is 4:1. Therefore, when helper functions are not present the level of splicing is lower, whereas when helper functions are present the level of splicing is higher (Mouw and Pintel, 2000). Mechanisms of mRNA processing further along the pathway are not currently understood.

Helper Functions

Adenovirus and herpes simplex virus are known to provide helper functions that create a productive infection for AAV-2. Adenovirus has four genes that are known to provide helper functions. Cellular protein YY1 represses the P5 promoter of AAV-2, however the adenovirus (1) E1a gene product relieves this repression and further activates transcription (Shi et al., 1991). The adenovirus (2) E2a gene product is a single-stranded DNA binding protein that increases the processivity of AAV-2 DNA replication (Ward, 1998). The adenovirus (3) E1 and (4) E4 gene products also possess helper functionality related to the production of double-stranded AAV-2 DNA (Fisher KJ, 1996). The E1 and E4 gene products act as a ubiquitin ligase and target the Mre11/Rad50/Nbs1 complex, which functions in DNA repair and inhibits AAV-2 transduction (Schwartz et al., 2007).

Herpes simplex virus is known to have four genes which provide helper functions. These four genes are a subset of the seven total herpes simplex virus replication genes. Herpes simplex virus genes (1) UL5, (2) UL6, and (3) UL52 encode the helicase primase complex (Weindler and Heilbronn, 1991). Herpex simplex virus gene (4) UL29 encodes the major ssDNA binding protein (Weindler and Heilbronn, 1991). The helper functions provided by herpes simplex virus are therefore suggested to provide mainly DNA replication functions for AAV-2 (Weindler and Heilbronn, 1991).

Rep Proteins

The two largest *rep* gene products are Rep 78 and Rep 68, which are differentiated by a splice site. Either one of these *rep* proteins are required for the replication of the AAV-2 genome (Ni et al., 1994; Ward et al., 1994). Rep 78 and Rep 68 have been shown to have several different enzymatic activities which function in DNA replication. Both *rep* proteins can bind to the 5' end of the AAV-2 ITR, which can fold into hairpin structures. The *rep* proteins (1) have site and sequence-specific endonuclease activity that nicks the AAV-2 DNA at the so-called terminal resolution site within the ITR (Im and Muzyczka, 1990). This allows the replication of the linear genome. Both *rep* proteins also (2) possess ATPase activity (Wonderling, Kyostio, and Owens, 1995), and (3) possess helicase activity (Wonderling, Kyostio, and Owens, 1995). Rep 78 may also interact with numerous cellular factors (Weitzman and Linden, 2011).

The two smaller *rep* proteins, Rep 52 and Rep 40 are also differentiated by a splice site. Neither *rep* proteins are required for AAV-2 replication or bind to the AAV-2 ITR. However, both Rep 52 and Rep 40 possess helicase activity and it is suggested that they are required for the accumulation and packaging of single stranded AAV-2 DNA (Weitzman and Linden, 2011). Rep 52 has also been shown to readily form complexes with viral capsid proteins (Dubielzig et al., 1999). To a lesser extent, Rep 78, Rep 68, and Rep 40 have also been observed to form complexes with viral proteins (Dubielzig et al., 1999).

The *rep* proteins also seem to play a role in the integration of the AAV-2 genome into chromosome 19. Specifically, AAV-2 integrates into the arm of chromosome 19 site-specifically at the target locus, called AAVS1 (Kotin, Linden, and Berns, 1992). When *rep* proteins are not present, such as in the case of recombinant AAV-2 vectors, it has been shown that AAV-2 genome integration is random (Walsh et al., 1992). It has been shown that integration functionality can be restored via Rep 78 (Balague C, 1997). Additionally, the *rep* binding site and the terminal resolution site have been mapped to the target sequence in chromosome 19 (Urcelay et al., 1995).

Assembly

Cap genes express VP proteins which rapidly assemble into empty capsids (Myers and Carter, 1980b). Capsid proteins must translocate into the nucleus before assembly. VP-1 and VP-2 possess a nuclear localization sequence, which allow for translocation across the cell nucleus (Hoque et al., 1999). However, while VP-3 is the major structural protein for AAV-2, it cannot generally translocate across the cell nucleus (Hoque et al., 1999). It has recently been reported that the expression of an activated assembly protein, which is nested in an alternative open reading frame within the *cap* gene, may be responsible for translocation of VP-3 into the nucleus (Sonntag, Schmidt, and Kleinschmidt, 2010). Capsids may assemble freely without the presence of *rep* proteins, although *rep* proteins are responsible for the redistribution of capsids outside the nucleus (Wistuba et al., 1997). Rep 52 and Rep 40 are not required for replication, but are required for the formation of full AAV-2 particles (Chejanovsky and Carter, 1989). Packaging of viral ssDNA is a relatively

slow process that is suggested to proceed in a 3' to 5' direction with the helicase activity of *rep* proteins driving encapsidation (King et al., 2001; Myers and Carter, 1980a). Packaging of the single-stranded AAV-2 genome occurs through a pore at the 5-fold axis where the *rep* proteins unwind the AAV-2 double-stranded DNA (Dubielzig et al., 1999; King et al., 2001).

AAV Serotypes

There are 12 AAV serotypes that have been proposed or established. A unique serotype is one that is antigenically distinct from other serotypes. AAV serotypes 2, 3, 5, and 6 natively infect humans, whereas serotypes 1, 4, 7-12 natively infect nonhuman primates (Gao, Vandenberghe, and Wilson, 2005). AAV-6 is considered a variant of AAV-1 due to a sequence identity of 96% (Gao et al., 2004). AAV-8 antiserum has been shown to block proposed AAV serotype 10 transduction. Similarly, AAV-4 antiserum was shown to block proposed AAV serotype 11 transduction. Therefore, it has been suggested that AAV-10 and AAV-11 do not warrant classification as distinct serotypes (Vandenberghe et al., 2006). However, it has also been reported that there is a much larger and diverse array of AAV forms, which may be segregated into clades according to serological similarities (Gao et al., 2004). Additionally, between many of the known AAV serotypes it has been suggested that there is a strong linkage of seroreactivity (Calcedo et al., 2009).

AAV sequences are more conserved in the *rep* region of the genome and possess more variation in the *cap* region (Gao, Vandenberghe, and Wilson, 2005). Relative to

AAV-2, similarity of the VP-1 sequence similarity ranges from 58% to 88% (Table 1). AAV-3b has the highest sequence similarity to AAV-2 in the capsid region, whereas AAV-5 has the lowest sequence similarity.

	Similarity to AAV-2 VP1	Host		Similarity to AAV-2 VP1	Host
AAV-1	83%	Simian	**AAV-7**	82%	Simian
AAV-2	100%	Human	**AAV-8**	83%	Simian
AAV-3b	88%	Human	**AAV-9**	82%	Simian
AAV-4	61%	Simian	**AAV-10**	84%	Simian
AAV-5	58%	Human	**AAV-11**	63%	Simian
AAV-6	83%	Human	**AAV-12**	60%	Simian

Table 1. Capsid sequence similarity of AAV serotypes 1-12 as compared to VP1 of AAV-2. Native host of the relative AAV serotype is also listed (Schmidt et al., 2008; Weitzman and Linden, 2011).

Tertiary Structure

The structure of the AAV-2 capsid has been modeled based on the VP-3 sequence. The N-terminal region of VP-1, VP-2, and the first 15 residues of VP-3 were not modeled in the crystallographic structure due to the disordered nature of those regions (Xie et al., 2002). However, those regions are expected to be present on the capsid interior and should not affect external capsid features (Kronenberg et al., 2005).

The structures between AAV serotypes are highly conserved and contain an antiparallel β-barrel and an α-helix (Chapman and Agbandje-McKenna, 2006). The β-barrel forms the contiguous capsid shell where the majority of the capsid surface is mostly composed of loop insertions between β-strands (Agbandje-McKenna and

Kleinschmidt, 2011). The interior of the capsid is formed by a four-stranded β-sheet from each VP-3 monomer. Surface exposed loops interweave near the 3-fold, where they create spike-like protrusions, near the 2-fold, where they create a depression, and near the 5-fold where they create a cylinder-like protrusion with an empty channel (Agbandje-McKenna and Kleinschmidt, 2011).

Topology

Adeno-associated virus 2 has several topological features of interest (Figure 3). The cylinder is a raised feature of AAV-2 that lies at the 5-fold symmetry axis. Surrounding the cylinder is a depression called the canyon. Near the 3-fold axis of symmetry there are three distinct raised features called spikes. Extending from each of the spikes is a moderately raised feature that is adjacent to the canyon. Along the 2-fold axes of symmetry there is a deep and narrow depression (Agbandje-McKenna and Kleinschmidt, 2011).

When the capsid sequences for AAV serotypes are compared to one another, it is apparent that there are nine segments of the VP-1 protein that are especially variable (Govindasamy et al., 2006). These protein segments which have higher variability have been dubbed variable region I (VR I) to variable region IX (VR IX) (Govindasamy et al., 2006). The variable sequences reside of the most exposed surfaces of AAV. As it relates to AAV-2, VR II consists of most of the cylinder. VR I, III, VI, and IX are present on the plateau. VR IV, V, VII, and VIII are present on the spike (Agbandje-McKenna and Kleinschmidt, 2011). The canyon and the 2-fold depression

are largely conserved and do not contain any of the aforementioned variable regions. The lack of variable regions and relatively less accessible surface along the canyon and 2-fold depression have led to speculation that these topological features may not play a primary role in determining the antigenicity of AAV-2 (Lochrie et al., 2006).

The HSPG receptor has been reported to be on the inside of the spike towards the 3-fold axis of symmetry (O'Donnell, Taylor, and Chapman, 2009). HSPG is the primary cell receptor for AAV-2 (Summerford and Samulski, 1998), suggesting that initial viral attachment occurs on the inside of the 3-fold spikes. Interaction with a secondary receptor may occur in another, possibly adjacent, location. Antibodies that bind to inside of the spike would therefore be suspected to impede initial attachment of AAV-2 to the cell membrane.

The VP1u region is located on the N-terminal end of VP-1. The region was not modeled in the crystallographic structure of AAV-2 (Xie et al., 2002), for two potential reasons that are not mutually exclusive. First, the region may be disordered. Second, VP-1 comprises approximately $1/10^{th}$ of a complete AAV-2 capsid, thus depleting the density of the region 10-fold when averaged icosahedrally. It has been reported that the N-terminus may have been visualized through *cryo*-EM as "fuzzy globules" on the inside of the capsid (Kronenberg, Kleinschmidt, and Bottcher, 2001). Additionally, it has been speculated that the N-terminal ends may be exposed through the cylinder before exiting the endosome

and may be important in endosomal escape and nuclear entry (Bleker, Sonntag, and Kleinschmidt, 2005).

The AAV-2 capsid maintains icosahedral symmetry and therefore every residue is repeated over the entire capsid 60-fold. General topological features within the asymmetric unit are repeated 60-fold over the capsid as well, including the plateau and each individual spike. The topological features called the cylinder and canyon, which are comprised of multiple subunits surrounding each 5-fold axis, are repeated 12-fold over the capsid surface. Similarly, the topology near each 3-fold consists of 3 spikes separated by valleys that meet over the 3-fold, where each 3 spike conglomerate is repeated 20-fold over the capsid surface.

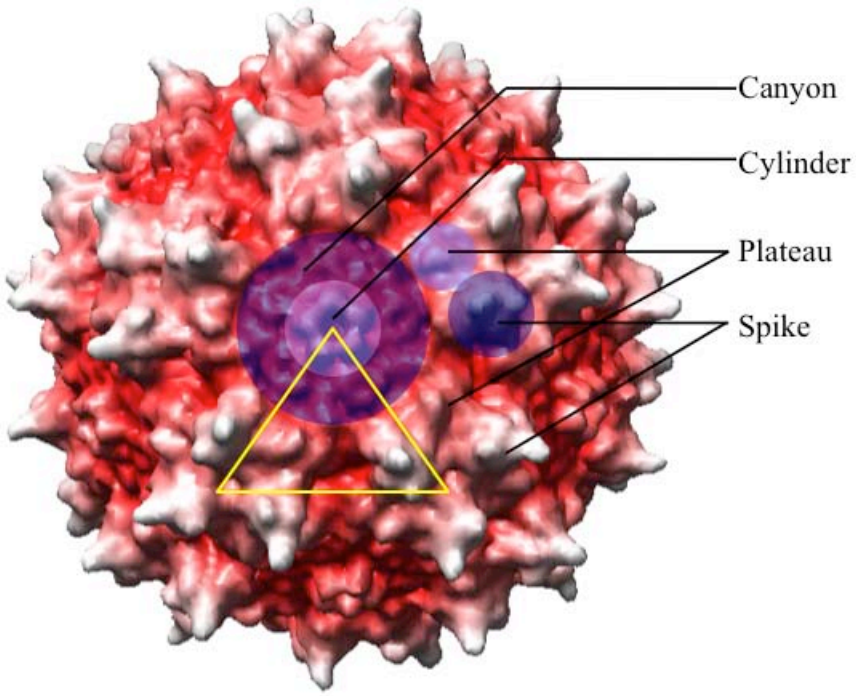


Figure 3. Topological features of AAV-2 are displayed on the crystallographic structure. The surface of AAV-2 has been simulated to 10 Å resolution using Chimera molecular imaging software (Pettersen et al., 2004). Major topological features of AAV-2 are overlaid in blue. The asymmetric unit of AAV-2 is outlined in yellow.

GENE THERAPY

Gene therapy involves the delivery to cells of DNA designed to alleviate disease of genetic nature or predisposition, DNA that commonly encodes a functional copy of a gene to replace a mutated form. This treatment requires a delivery vehicle to transport therapeutic DNA into a target cell which is genetically deficient, a role which is particularly suited to viruses (Hodgson, 1995). Currently there are no

licensed gene therapy treatments in the U.S. However, human studies with gene therapy vectors have shown successful results with both integrating and non-integrating viruses (Mingozzi and High, 2011; Naldini, 2011).

There are several candidates for gene therapy vectors, including adenoviruses, retroviruses, and adeno-associated viruses (Pfeifer and Verma, 2001). Retroviral vectors have generally been limited to dividing cells and integrate randomly into the host genome, carrying risk of oncogenic transformation (Flotte, 2007). Oncogenic transformation from retroviral vectors has been shown to cause leukemia in some patients due to integration of the vector into the LMO-2 oncogene (Cole, 2008). Adenoviruses do not integrate into the host genome, which causes transient expression of the transduced gene, and may be highly immunogenic, leading to dangerous and occasionally fatal systemic immune responses (Marshall, 1999).

AAV in Gene Therapy

AAV has become a leading candidate for gene therapy and is in use in current or recent gene therapy trials to target hemophilia B, prostate and melanoma cancers, Canavan disease, Alzheimer's, Parkinson's, muscular dystrophy, rheumatoid arthritis and HIV vaccines (Carter, 2005). Several factors contribute to the preferential use of AAV as a gene therapy vector. Immune responses to AAV are mild and non-inflammatory, significantly reducing the possibility of complications during treatment (Bueler, 1999; Chirmule et al., 1999). AAV is non-pathogenic and has not been associated with any pathogenicity beyond that of the infecting helper virus

(Berns and Linden, 1995). The non-pathogenicity of AAV is particularly favorable because contaminant wild-type virus with the therapeutic vector will not cause disease in the treated patient. AAV has been shown to integrate site-specifically into the arm of chromosome 19, reducing the risk of oncogenic transformation (Dutheil et al., 2000). However, the ITRs responsible for site-specific integration are often removed when producing the gene delivery vehicle. AAV has wide cell tropism and infects dividing and non-dividing cells, making it possible to transduce a large array of possible cell types (Carter, Burstein, and Peluso, 2008). AAV is also producible in large quantity, which makes it an efficient vector to produce for gene therapy treatments (Xie et al., 2004).

AAV also has several disadvantages compared to other gene therapy vector candidates. AAV's has a limited gene payload of 5 kb, limiting the gene size which may be used in therapeutic treatments (Dong, Fan, and Frizzell, 1996). If the gene payload is larger than 5 kb, then multipartite AAV vectors must be used (Zhang, Cortez, and Berns, 2007). Additionally, clinical trials involving experimental cystic fibrosis and hemophilia treatments required re-administration of therapies (Flotte and Carter, 1998; Kay et al., 2000). As only ~10% of progenitor cells below the lung epithelium are transduced, transgene expression falls over 6 months as surface cells are turned over (Carter and Flotte, 1996). This suggests that repeated administrations of gene therapy vectors may be required. However, exposure to natural AAV or vectors in animals and humans results in development of neutralizing antibodies (Halbert et al., 2000; Moskalenko et al., 2000; Peden et al., 2004; Xiao et al., 1999). This can decrease the efficiency of transgene expression

upon subsequent administration(s) of vectors of the same serotype (Davidoff et al., 2005; Glas and van der Linden, 2010; Manno et al., 2006; Wang et al., 2011). Additionally, immune response still presents a significant barrier to efficient delivery of the vector during initial treatment, as neutralizing antibodies have been found in 35-80% of the U.S. population (Calcedo et al., 2009).

Cellular immune response is largely humoral and directed against the viral capsid, while immune response to the transgene product is rare (Brockstedt et al., 1999; Chirmule et al., 2000; Halbert et al., 1997; Hernandez et al., 1999; Manning et al., 1998). It has been shown that modification of the AAV viral capsid may lead to reduced antibody binding and neutralization, with as little as one amino acid mutation (Lochrie et al., 2006). Therefore, it is widely thought that engineering antigenic variant vectors will have substantial impact in the development of efficient gene therapy treatments (Flotte, 2005; Peden et al., 2004; Xiao et al., 1997).

Escape Mutants

Studies of different virus families suggest that the AAV capsid surface may contain 2 to 3 immunodominant regions (Zuffi et al., 2001). Each structural study involving a monoclonal antibody will help to elucidate one immunodominant region per structure. It would be beneficial to have multiple structures per immunodominant epitope to fully determine the amino acids most targeted on the viral surface. A recent study of 8 different monoclonal antibodies bound to CPV reported that the footprint of the antibodies cumulatively covered more than 60% of the viral surface

(Hafenstein et al., 2009). Antibodies were reported to bind to one of two general immunogenic sites. At each immunogenic site, only several amino acids were common to the footprint of all antibodies. Determining the most immunodominant amino acids would be beneficial in future rational design of viral escape mutants as each immunodominant region could be altered with as little as one single-site mutation to create an antigenically distinct serotype.

Antibodies Against AAV-2

Several monoclonal antibodies have been isolated which are specific to AAV. Monoclonal antibodies C24-B, C37-B, and A20 all specifically bind to and neutralize AAV-2 (Wobus et al., 2000). C24-B and C37-B both inhibit binding of AAV-2 to host cells. Specificially, C37-B has been implicated in binding at or near VR V and VIII via PEPSCAN (Wobus et al., 2000). A20 is a monoclonal antibody that neutralizes AAV-2 and AAV-3 subsequent to primary receptor binding (Wobus et al., 2000). It is the most widely studied monoclonal antibody against AAV-2, due in part to its ability to bind only to fully assembled capsid and not to unassembled capsid proteins (Moskalenko et al., 2000; Wistuba et al., 1995; Wobus et al., 2000). A20 has been implicated in binding at or near VR I, III, VII, and IX via PEPSCAN (Figure 4) (Wobus et al., 2000). VR I, IV, and IX were also implicated via single-site mutagenesis (Lochrie et al., 2006).

It has been suggested that there are at least three neutralizing immunogenic sites on the capsid of AAV-2 (Lochrie et al., 2006), therefore it is unlikely that mutations

solely within the A20 footprint will yield a capsid that is completely distinct antigenically. However, two mutations that inhibit A20 neutralization also confer resistance to polyclonal neutralization, suggesting that A20 binding might model neutralization at a dominant epitope (Lochrie et al., 2006). There have been multiple attempts to locate the A20 epitope using various techniques, including PEPSCAN, peptide competition, peptide insertions, and both site-directed and scanning mutagenesis (Girod et al., 1999; Lochrie et al., 2006; Moskalenko et al., 2000; Shi, Arnold, and Bartlett, 2001; Wobus et al., 2000; Wu et al., 2000). Lochrie et al. (Lochrie et al., 2006) have drawn attention to inconsistencies between some of the experimental epitope mappings, and between some of the proposed immunogenic sites and the subsequent AAV-2 atomic structure (Xie et al., 2002), noting that inaccuracies result from known limitations of standard methods (Van Regenmortel, 1992). There has long been an interest in resolving these questions through direct visualization of an AAV-2/A20 complex (Lochrie et al., 2006).

Additionally, several groups have used polyclonal populations of antibodies to determine which regions of the AAV-2 surface are involved in general antigenicity. Reports from these studies have suggested that neutralizing antibodies from intravenous immunoglobulins (IVIG) samples interact with AAV-2 at VR I, III-IX (Figure 4) (Lochrie et al., 2006; Wu et al., 2000; Zaiss and Muruve, 2008).

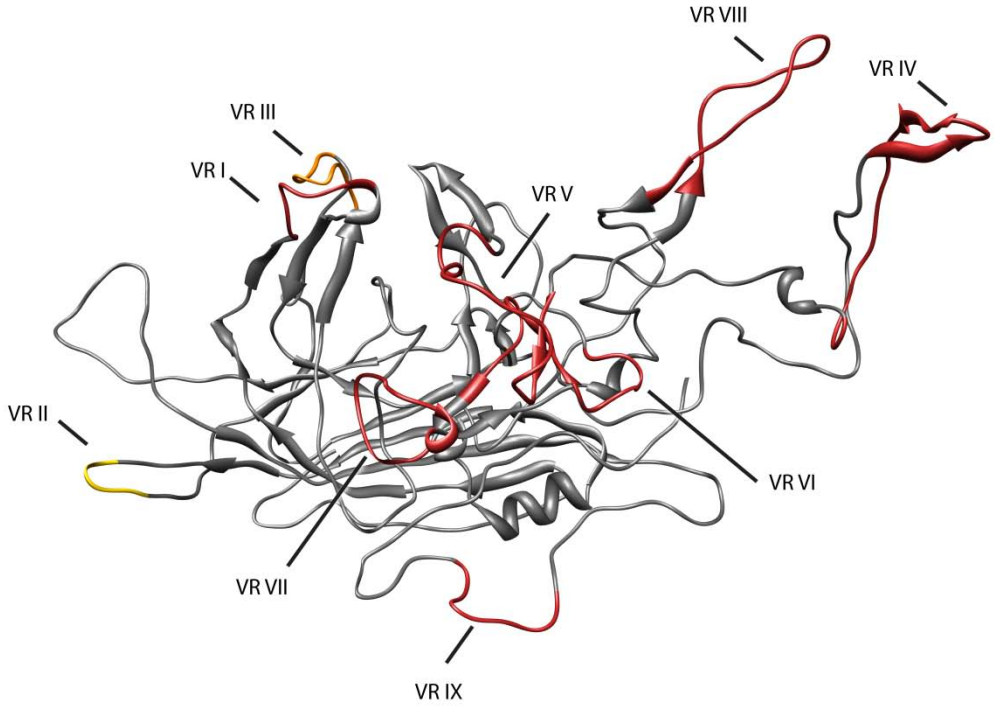


Figure 4. VP-3 subunit of AAV-2. Variable regions implicated in neutralization by polyclonal sera are colored in red (VR I, IV-IX). Variable regions implicated in neutralization by monoclonal antibodies only are colored in orange (VR III). Variable regions not implicated in neutralization by IVIG or monoclonal antibodies are colored in yellow (VR II).

***CRYO*-ELECTRON MICROSCOPY**

Cryo-EM has several distinctions as a structural approach to the assessment of large and possibly heterogenous macromolecular assemblies. *Cryo*-EM has a relatively expedient turnover time for the structure determination of molecules, although the technique does not yet match the resolution achieved in X-ray crystallography.

Cyro-EM has several advantages for the direct visualization of a complex of AAV-2 and A20 antibody. First, it does not rely on the crystallization of the complex, which is often a rate-limiting step. Second, *cryo*-EM visualizes the complex under native conditions and therefore a biologically relevant environment. Finally, high resolution is not a necessity because the effective resolution can be increased by fitting the high-resolution crystallographic structure of AAV-2 and the structure of A20 into a lower resolution electron density map of the complex. Fitting crystallographic data into an electron density map acquired through *cryo*-EM may increase the effective resolution of the reconstruction 5-fold (Fabiola and Chapman, 2005).

There have been several recent structural studies of antibodies complexed with members of the family *Parvoviridae* and *Picornaviridae* through *cryo*-electron microscopy. Human rhinovirus was complexed with monoclonal antibody MAb17-IA was imaged through *cryo*-electron microscopy and reconstructed to a resolution of 28 Å, where the antibody was reported to bind bivalently across the 2-fold symmetry axis where it is mediated by several electrostatic interactions (Smith et al., 1993). More recently, minute virus of mice was imaged in complex with monoclonal antibody B7 and reconstructed to a resolution of 7 Å, where the antibody was reported to bind monovalently to the 3-fold symmetry axis (Kaufmann et al., 2007). Canine parvovirus and feline parvovirus were imaged with 8 different monoclonal antibodies and reconstructed to a resolution of 8.5 Å to 18 Å, where the antibodies were found to bind monovalently and two major antigenic epitopes were determined from the cumulative overlap each individual antibody footprint

(Hafenstein et al., 2009). Additionally, the first structure of AAV complexed with its primary cellular receptor has also been determined to a resolution of 8.3 Å, where heparin was reported to bind near the 3-fold axis of symmetry (O'Donnell, Chapman, and Taylor, 2008).

The analysis of reconstructions acquired through *cryo*-EM has been largely reliant on either the crystallographic structures of the individual proteins being previously determined or the availability of homologous structures where a homology model may be determined. In each previous case, the crystallographic model of the respective virus was used in the analysis of the electron density map of the complex. Additionally, the antibody structure used in the analysis of each complex was either determined through crystallography or a homology model was used.

Single-Particle Cryo-electron Microscopy

Single-particle electron microscopy is an imaging and data analysis technique that is used to determine the structure of single-particle images. Single-particle methods rely on imaging identical many similar particles in a distribution of orientations, rather than on imaging the same particle in many different orientations—such as in electron tomography. It is particularly useful for relatively large biological macromolecules, particularly those exhibiting high symmetry. However, the particles imaged must be very similar or identical in structure. Particles should also be imaged in a relatively uniform distribution of orientations or the acquired data will be less robust for poorly sampled orientations. Ultimately, the population of

images acquired from single-particle electron microscopy may be reconstructed into a 3-D model.

EMAN is single-particle reconstruction software suite for performing single-particle reconstructions. The methodology used by EMAN allows it to rapidly converge on a final reconstruction, which is imperative when using very large data sets. Therefore, data analyses for single-particle *cryo*-EM will be approached with respect to EMAN.

Data Acquisition

Cryo conditions are necessary for visualizing a complex in its native state without stain or fixation. Under cyro conditions the sample is flash-frozen, creating an environment where the particles are embedded in a thin layer of amorphous ice that may be later removed through averaging techniques. The remaining signal represents the sample particle itself.

Often sample vitrification occurs on holey carbon grids, where the sample is imaged through fabricated holes in the carbon layer. Specimens are applied to EM grids and excess sample is removed immediately prior to vitrification. Thick sections of vitrified ice cause complications to arise in particle alignment, contrast transfer function correction, and create increased electron scatter from disordered ice. Similarly, sections of vitrified ice that are disproportionately thin may flatten particles, significantly reduce particle concentration around hole centers, and cause salt crystallization due to relative high evaporation rate for small solution volumes. Freezing on the order of 10^6 degrees per second is required to vitrify amorphous

ice, which reduces ordered noise. Ordered cubic or hexagonal ice form at lower freezing rates or when the sample begins to thaw, both of which impair the ability to remove ice signal through the averaging of successive images.

Sample preparation is empirical and not always consistent, although recent technological advances have automated a majority of the process. The Vitrobot (FEI) is seeing increased use for sample preparation in *cryo*-EM. A sample is placed into the Vitrobot's main chamber where a specific humidity is set. Through an automated process, the sample is blotted for a specific time and the sample is freeze-plunged. However, even with automation the process is not entirely consistent. Thus, multiple samples are often prepared in order to mitigate the remaining inherent variation in preparing otherwise identical samples.

Data collection is most commonplace a charge-coupled device (CCD), although film is still in use. For data acquisition where a CCD is utilized, automated data collection methods may be used to increase the volume of data collected by using a software suite such as the increasingly popular Appion (Lander et al., 2009). Unfortunately, due to the digital nature of a CCD, it is possible for the resolution of a reconstruction to be limited by the pixel size used in data acquisition. It order to avoid this, the pixel size must be no greater than one-half the expected resolution of the final reconstruction.

The Nyquist-Shannon sampling theorem states that the exact reconstruction of a continuous-time baseband signal from its samples is possible if the signal is bandlimited and the sampling frequency is greater than twice the signal bandwidth.

Therefore, for bandlimited signals, the Nyquist frequency is one-half the sampling frequency of a discrete signal. It is also equivalent to the pixel size that must be used to unambiguously reconstruct a model of that resolution. However, bandlimited signals are theoretical constructs that are not feasible to create under real-world conditions, making oversampling essential if the target resolution is to be achieved.

Requirements Necessary for Particle Reconstruction

Images acquired from *cryo*-EM are two-dimensional projections of the particle, many of which are combined in iterative 3-D reconstruction methods to create a representative 3-D electron density map. Icosahedral particles are predisposed to being present in a large number of orientations on any particular grid. Icosahedral particle reconstruction takes advantage of the large orientation sampling, as well as high particle symmetry and the ability to acquire large data sets, to procure the highest resolution reconstruction possible.

There are three different requirements that must be met in order to accurately reconstruct an electron density map. These requirements include selecting sample particles, creating an initial electron density map, and determining relative amplitude correction parameters.

Particle Selection

Particle selection involves isolating individual particles which will be used as the primary data set in the reconstruction process. Particles must be imaged and the

particles must subsequently identified. Identification of particles is approximately 80% automated using a software suite such as Appion (Lander et al., 2009). Appion uses a combination of automation and user input to crop particle images from digital electron micrograph images into an array that will be used for processing. Approximately 10% of particles are incorrectly selected and another 10% are unselected by the automated process, requiring manual intervention to acquire the most robust data set. However, manual intervention may not yield a significant increase in resolution when a large data pool (greater than 10,000 particles) is acquired, as averaging routines dampen the impact of incorrect particle selection. Images are generally cropped at approximately 115% the diameter of the particle. Particle images cropped too close (1) increase the chance of cropping the particle itself during subsequent transformations during class averaging routines and (2) may introduce Fourier artifacts from the edge of the cropped image. Particle images cropped too far from the particle have an increased risk of incorrect particle alignment during later stages of the reconstruction process.

Initial Electron Density Map

An initial electron density map is a requirement of the EMAN reconstruction routines. It is not necessary to have a precise, or even similar, initial electron density map, but a course approximation of the particle's structure is necessary for initial reconstruction routines. When a derivative of a known structure is used as an initial electron density map, the refinement process may be biased in favor of the initial model. Therefore, it is best to use model-independent methods to begin refinement.

In many cases, an initial electron density map may be created directly from the original data set.

EMAN provides routines for initial model generation that involve searching for projections with the best two-fold, three-fold, and five-fold symmetries (Ludtke, Baldwin, and Chiu, 1999). The dot product of the projection and itself with the appropriate symmetry applied is used to determine the quality factor for the particle in the respective orientation. For example, if a particle image lies on a 3-fold axis of symmetry, then the particle image would not substantially change when rotated 120°. Therefore, the quality factor would be high when the dot product of the original image and the image with a 120° rotation applied was calculated. Particles with a high quality factor for a particular orientation are deemed to be in said orientation. After the particles with the best symmetries are determined, the class averages of those particles are used to create the initial electron density map.

Amplitude Correction

Amplification correction involves correcting the particle images for systematic distortions caused by the electron microscope. Generally, a transfer function is a mathematical descriptor of the relationship between the input and output of a system. For electron microscopy, the contrast transfer function (CTF) is the descriptor of the relationship between the contrast of the real particle and the contrast of the acquired image. We may consider an image acquired from electron microscopy as a convolution of the image contrast and a point spread function,

which modulates the image contrast in a spacial-frequency dependent manner. The CTF is the Fourier transform of the point spread function, which is the correction that may be applied in order to deconvolute the original image.

In an ideal situation, the intensity of the electron microscope's beam would be constant over all wavelength frequencies. However, this theoretical situation is not encountered under real-world conditions. In order to acquire optimal resolution, corrections must be applied to acquired data in order to account for variations in the beam's intensity over various wavelength frequencies.

Therefore, the contrast transfer function must first be determined for the particle set. EMAN uses the following formula to approximate the CTF:

$$\text{I.} \quad C(s) = A\left[\sqrt{1 - C_A^2}\, sin(\gamma) + C_a\, cos(\gamma)\right]$$

$$Where\ \gamma = 2\pi\left(\frac{C_S \lambda s^4}{4} + \frac{\Delta Z \lambda s^2}{2}\right), 0 \leq C_A \leq 1$$

Where C_S is the spherical aberration of the microscope, ΔZ is the distance from focus, s is the distance from the particle origin, A is the amplitude of the CTF, and λ is the wavelength corresponding to the acceleration voltage of the microscope. This formula illustrates the dependence of the CTF on defocus (ΔZ), as C_S and λ are constant functions of the microscope. An EM grid is never entirely flat or level, which causes small relative changes in the height of each image. Each micrograph will therefore be at a slightly different defocus and, in turn, the CTF will vary with each image.

Practical determination of the CTF involves sampling high-signal regions of an EM grid such as from the carbon film or from the particles themselves. From these regions, a radially averaged power spectrum may be determined. EMAN uses the following formula to approximate the radially averaged power spectrum:

$$\text{II.} \quad M(\phi,\sigma,s)^2 = C(s)^2\, E(s)^2\, F(\phi,\sigma,s)^2 + N(s)^2$$

Where $M(\phi,\sigma,s)$ is a measured Fourier amplitude of the projection, $C(s)$ is the contrast transfer function, $E(s)^2$ is the envelope function, $F(\phi,\sigma,s)$ is the structure factor of the particle, and $N(s)$ is noise. The noise is determined empirically from each individual micrograph and incorporates various physical effects such as incoherent scattering, film noise, and scatter noise. $N(s)^2$ may be subtracted from $M(\phi,\sigma,s)^2$, and the formula for the radially averaged power spectra becomes:

$$\text{III.} \quad M'(\phi,\sigma,s)^2 = C(s)^2\, E(s)^2\, F(\phi,\sigma,s)^2$$

Where $M'(\phi,\sigma,s)^2$ is a measured Fourier amplitude of the projection corrected for noise. Determination of the corrected amplitudes relies on the simple algebraic adjustment:

$$\text{IV.} \quad F(\phi,\sigma,s) = \frac{M'(\phi,\sigma,s)}{C(s)\, E(s)}$$

The CTF and envelope function can be empirically determined from radially averaged power spectrum. It is tempting to determine the structure factor by dividing the measured amplitude of the projection by the CTF and envelope function. However, when the CTF approaches zero no signal is acquired, thus the

correction approaches infinity. Therefore, data near the CTF nodes may be excluded or a filter may be applied to limit the unnecessary inclusion of high levels of noise in the structure factor at spacial frequencies where the CTF is small. A Wiener filter is the CTF is adjusted by:

$$\text{V.} \qquad W(s) = \frac{C(s)}{C(s)^2 + \frac{1}{SNR}}$$

Where SNR is the signal to noise ratio and W(s) is the Wiener filter. Applying the Weiner filter to equation IV gives:

$$\text{VI.} \qquad F(\phi, \sigma, s) = \frac{M'(\phi,\sigma,s)\, W(s)}{E(s)}$$

For the case where the SNR approaches zero, the Wiener filter will also approach zero, which is sensible because the estimate of the corrected amplitudes (or structure factor) should be zero when no signal is present. For the case where the SNR approaches infinity, the Wiener filter will approach the inverse of the CTF, and equation VI will be equivalent to equation IV. In intermediate cases where the CTF is small, the Wiener filter applies a conservative restoration of signal that would be lost in equation IV. Thus, the Wiener filter mitigates artifacts in cases where the CTF is small or zero.

Particle Reconstruction

Particle reconstruction is the largely automated process of creating a 3-D electron density map from the original particle images. Particle images collected on a CCD

are effectively 2-D projections of the 3-D particle which was originally imaged. It is incumbent on reconstruction routines to take the 2-D projections and recreate a 3-D electron density map of the original particle. Due to the large amount of data that must be analyzed in most reconstructions, it is computationally efficient to represent the particle images as the Fourier transforms of the 2-dimensional particles. Representing the particle as its Fourier transform also allows for the efficient correction of amplitudes based on the CTF, envelope function, and noise. All steps which involve amplitude corrections are therefore performed in Fourier space. EMAN also uses a Fourier space reconstruction routine to construct the final 3-D model. A flow-chart diagram of the reconstruction process is diagramed in **Error! Reference source not found.**.

Reconstruction Routines

Once the necessary initial parameters have been determined for the data set, an environment must be created where particle images may be classified according to the orientation of the initial electron density map. Initial projections are created from a preliminary map to produce a reference for classification of particle images. Each projection represents the particle in a different orientation. Separately, each particle image is amplitude corrected according to the CTF, envelope function, and noise.

Particle classification and class averaging is the bulk of the refinement process. In summary, particles are classified according to an orientation and all particles of a

particular class are averaged to increase the SNR. To this effect, each particle projection is rotationally and translationally aligned to each reference class, where the correlation is determined by the dot product of the two images. Each particle is grouped according to which reference class it is most highly correlated to. Particles within each individual class are translationally and rotationally aligned and projections with a poor fit to the class are removed. The aligned projections are then averaged to create a new projection for that class. Averaged classes reduce uncorrelated noise, which roughly corresponds to a factor of $\sqrt{N}$, where N is the number of images. The new class averages are then amplitude corrected and assigned as a new reference class for the next iteration of class averaging. This process is repeated new class averages converge with the reference classes.

Once class averages have converged, a 3-D electron density map may be created. The orientation of the class averages may be different from the original reference classes. Therefore, the class averages are amplitude corrected and assigned a new orientation (specified by Euler angles), relative to the original model. Averaged projections from each class are back-projected to reconstruct an icosahedral electron density map of the particle, which embodies all orientation classes. The reconstruction may be iterated with the newly reconstructed map, refined against a larger or more refined data set, or accepted as a final model if objectives have been met.

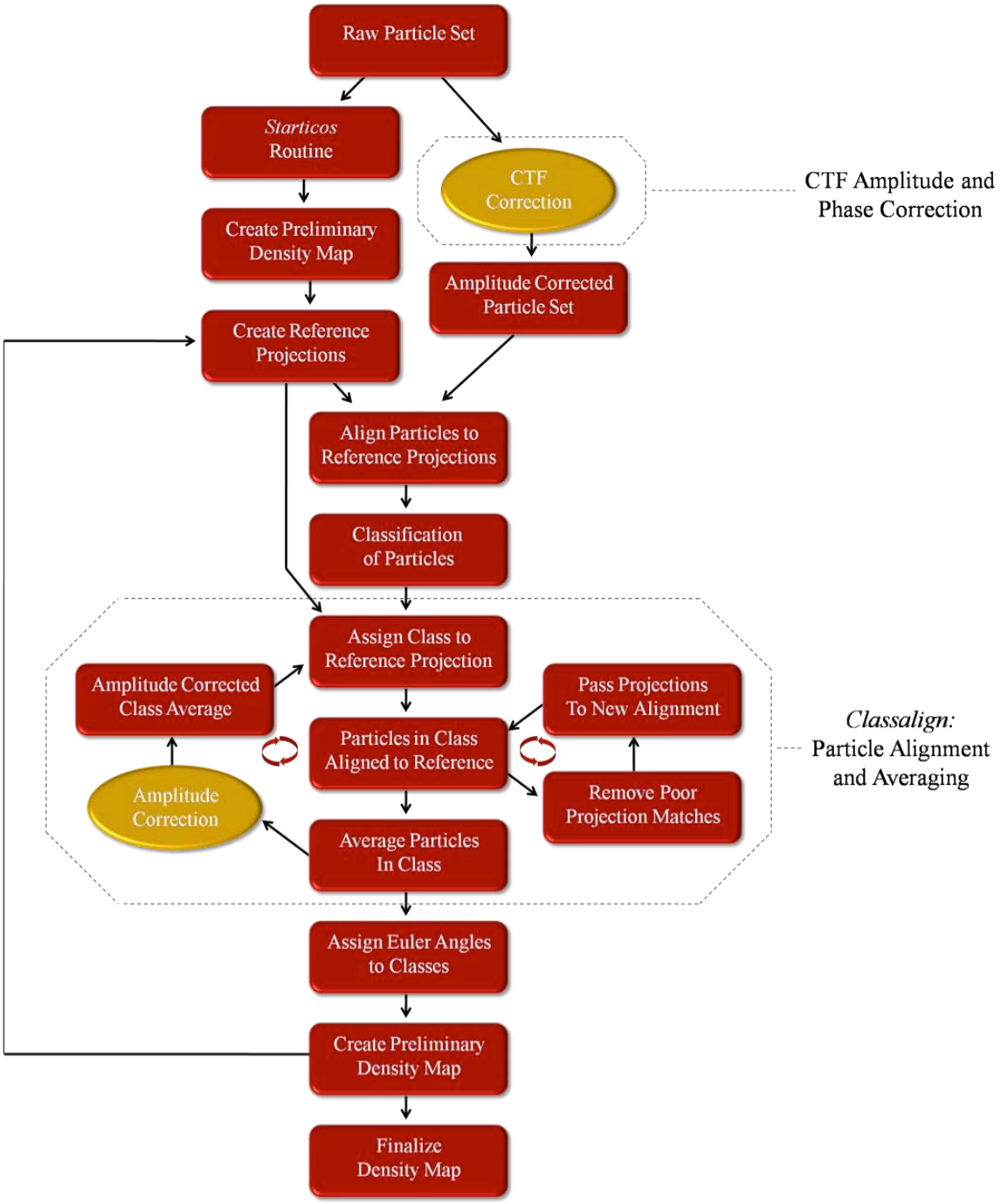


Figure 5. EMAN reconstruction routines based on methods described by Ludtke et al. (Ludtke, Baldwin, and Chiu, 1999).

ROADMAP

Initial attempts at producing conventional Fab are discussed in "Preparation of MAb A20 and Fab Fragments". The purification of adeno-associated virus 2 and A20 monoclonal antibody is addressed. The A20 antibody was functionally assessed for its capability to bind to AAV-2. A20 Fab was prepared and the functionality is assessed through several various means. In "Electron Microscopy of AAV-2 Complexed with A20 MAb", electron microscopy was used to image native A20 antibody bound to AAV-2 on two different prototypes of functionalized EM grids.

The manuscript describing the binding site of A20 Fab' on AAV-2 is discussed in the section "Structure of Adeno-associated Virus-2 in Complex with Neutralizing Monoclonal Antibody A20". In this section, A20 Fab' was prepared and imaged with AAV-2 using single-particle *cyro*-EM and an electron density map was reconstructed. Electron density maps were magnification corrected based on crystallographic models. A homology model was created based on the sequence of A20 and the quality was assessed against crystallographic models of similar sequence. The homology was fitted into the newly-created electron density map and the binding site was determined.

Further unpublished analysis of A20 Fab' is discussed in "Afterword: AAV-2 in Complex with Anitbody A20". In this section, the functional analysis of A20 Fab' is assessed, followed by an analysis of the electrostatics and atomic solvation parameters of the A20 binding site for binding and non-binding AAV serotypes.

A summary of the results learned here and proposed future direction for this research is discussed in “Review and Future Direction”.

This report concludes with section “Appendix I: Creation of AAV-2 Single Site Mutant Plasmids”, where an alternative approach is taken to determining the A20 epitope is discussed. Here an E. coli strain that is deficient in its DNA repair pathway is used to create single-site mutations on the AAV-2 plasmid.

PREPARATION OF MAB A20 AND FAB FRAGMENTS

DISCLAIMER

This chapter includes preliminary investigations that did not contribute directly to the structure of the AAV-2/A20 complex presented in chapter 4. A20 Fab, as prepared here, did not bind strongly to AAV-2. The sections "Production and Purification of Adeno-Associated Virus 2", "Production and Purification of A20 Monoclonal Antibody", and "Assessment of A20 MAb Functionality" provide details of preparatory techniques and analysis which were foundations for investigations described in later chapters where they will be covered only in summary form.

BACKGROUND

Antibodies of the IgG isotype consist of four protein chains, including two heavy chains and two light chains. The chains are named according to their size, where the heavy chains are approximately 50 kD and two light chains are about 25 kD. Each antibody consists of two functional Fab domains connected to a functional Fc domain through a hinge region. Each Fab domain consists of a light chain and one-half of the heavy chain, where the two remaining halves of the heavy chain form the Fc domain. Each domain of the antibody consists of two β-barrels contributed from each chain, for a total of four β-barrels per each domain.

The Fab domains contain the antigen-binding regions of the antibody and the Fc domains bind to Fc receptors on immune cells. IgG antibodies are bivalent due to containing two antigen-binding Fab domains. The bivalent nature of antibodies allows for the crosslinking of an antibody to two antigens, which causes aggregation. For structural assessments of IgGs bound to viral antigens, precedence has been to cleave the Fab domains of the antibody to create a monovalent domain with a single paratope (Hafenstein et al., 2009; Kaufmann et al., 2007; Smith et al., 1993). The cleavage of the Fab domains creates a monovalent domain from the bivalent antibody.

Structure determination of the binding site of monoclonal antibodies on viruses has been performed several times before. To the best of our knowledge, no complexes of the Dependovirus genus have yet been determined. However, Fab-virus complexes of human rhinovirus (HRV), Minute virus of mice (MVM), canine parvovirus (CPV), and feline panleukopenia virus (FPV), have been determined through single-particle cyro-EM (Hafenstein et al., 2009; Kaufmann et al., 2007; Smith et al., 1993).

For the HRV structure, the crystallographic structure of both the virus and Fab-17-IA were determined and fitted into a reconstruction of the antigen-Fab complex, which was determined via *cryo*-EM (Smith et al., 1993). For MVM, the structure of the virus was determined, a homology model of neutralizing Fab B7 was generated, and a reconstruction of the respective complex was acquired via *cryo*-EM. The homology model of B7 was then fitted into the reconstructed complex (Kaufmann et al., 2007). For CPV, the crystal structure of the virus and one Fab was determined.

Complexes of CPV and FPV and 8 different Fabs of neutralizing MAbs were reconstructed by *cryo*-EM. For the 7 Fabs where crystal structures were not available, homology models were used (Hafenstein et al., 2009).

To summarize the approach used in all of these prior studies, Fab fragments were purified for the preparation of antigen-Fab complexes. The complex structures were then determined via *cryo*-EM. Either the crystal structure of the Fab or a homology model was used in the determination of the final complex structure. In each case, the crystal structure of the virus was already known.

In this section, we attempt to prepare functional Fab domains of monoclonal antibody A20 that bind to AAV-2. The crystal structure of AAV-2 was previously determined (Xie et al., 2002). The preparation of monovalent Fab would allow for the imaging of AAV-2 complexed with Fab and the subsequent determination of said structure. If successful, either the crystal structure of the Fab could be determined or a homology model of A20 could be generated. The Fab model could then be fitted into the complex density yielding a structure of the antigen-Fab complex.

PRODUCTION AND PURIFICATION OF ADENO-ASSOCIATED VIRUS 2

Methods

AAV-2 was propagated in HeLa cells by methods previously described (Xie et al., 2004) and summarized here. Cells were thawed from frozen stock and were grown in Joklik's Modification of Eagle's Minimal Essential Medium (JMEM) (Sigma) with

10% Cosmic Calf Serum (CCS) (HyClone) and 10 mg/L of antibiotic-antimiotic containing 100 units/mL of penicillin, 50 mg/L streptomycin, 25 mg/L of amphotericin (Gibco), and 10 mg/L of gentamicin (Invitrogen). Cells were initiated in T25 flasks, transferred to T75 flasks, then transferred to T225 flasks.

AAV-2 propagation was completed in two steps, including a primary transfection and a secondary infection. For the primary transfection pAV2 plasmid was kindly provided by K. Berns, which is a pBluescript plasmid that contains the full AAV-2 genetic sequence (Laughlin et al., 1983). Adenovirus 2 was kindly provided by E. Falck-Pederson. Hela cells at 80% confluence were transfected with pAV2 at 37°C and then additionally transfected with Adenovirus 2 four hours later. HeLa cells were then incubated 48 hours at 37°C. The cell lysate was used as primary inoculum for the secondary transfection.

Inoculum from the primary transfection was further prepared by adding 14% of the transfection product to 2% CCS. The inoculum was added to HeLa culture at 80% confluence and incubated for 4 days at 37°C. Infected HeLa cells were freeze-thawed three times and then purified by three cesium chloride density gradient ultracentrifugations for 24 hours each in a SW55 swinging-bucket rotor at 38,000 rpm. Purified samples were dialyzed into 100 mM HEPES, 50 mM magnesium chloride, and 5% glycerol (5% HMG) in order to maintain solubility.

Results

Purified AAV-2 was assessed by protein gel electrophoresis of denatured AAV-2 capsid protein (Figure 6). AAV-2 contains VP1 (87 kD), VP2 (73 kD), and VP3 (62 kD) in a 1:1:10 ratio, respectively. Protein band intensities of denatured AAV-2 capsids are consistent with the molecular weights and relative abundance of the individual capsid proteins.

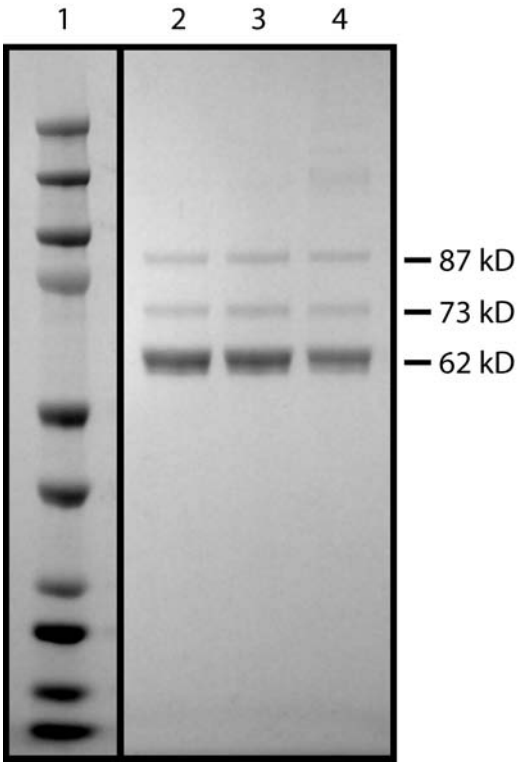


Figure 6. Protein gel electrophoresis of CsCl purified, denatured AAV-2. Protein was run against Kaleidoscope Precision Plus Protein Standard from Bio-Rad (lane 1). CsCl fractions are shown from immediately above the main band (lane 2), the main band (lane 3), and immediate below the main band (lane 4).

PRODUCTION AND PURIFICATION OF A20 MONOCLONAL ANTIBODY

Methods

A CELLine CL 1000 bioreactor was used to for its ability to produce a large amount of concentrated antibody in a compact system. Stocks of hybridoma cells for A20 monoclonal antibody were generously provided by Jürgen Kleinschmidt. $25x10^6$ hybridoma cells were seeded into the bioreactor and cultured for 7 days before harvesting. The bioreactor utilizes two compartments: a cell compartment which contained 10% FCS in RPMI, and a media compartment which contained 2% FCS in RPMI. After harvesting, cells were removed by centrifugation at 1,000 relative centrifugal force (RCF) for 10 min and cellular debris was removed via a 45 μm filter.

Affinity chromatography was used to purify A20 monoclonal antibody from the crude hybridoma supernate. Approximately 10 mL of crude supernate was passed through a 1 mL protein G column (GE Healthcare) at room temperature and washed with 20 mM sodium phosphate at pH 7.0. Subsequently, A20 bound to the column was eluted with 0.15 M sodium chloride and 0.1 M sodium acetate at pH 4.4 and immediately neutralized to pH 7.2 with 1 M Tris-HCl at pH 9.0. Elution fractions were collected in 2 mL volumes.

Results

Initial purification attempts revealed that purified A20 precipitates at concentrations higher than 1.0 – 2.0 mg/mL at 4°C. At high-concentrations some monoclonal antibodies have been reported to precipitate at low temperatures due to the self-association of Fc domains, thus earning the name "cryoglobulins" (Izui et al., 1993). Preparatory steps required the use of solubilized A20. Dissolving precipitated A20 was not feasible. Therefore, it was necessary to perform all experiments at room temperature when working at concentrations above 1.0 mg/mL.

Purified A20 monoclonal antibody gels showed a relatively pure band at 150 kD (Figure 7). Faint bands were observed at 125 kD, 100 kD, 75 kD, 50 kD, and 25 kD, consistent with a small amount of disulfide bond dissociation.

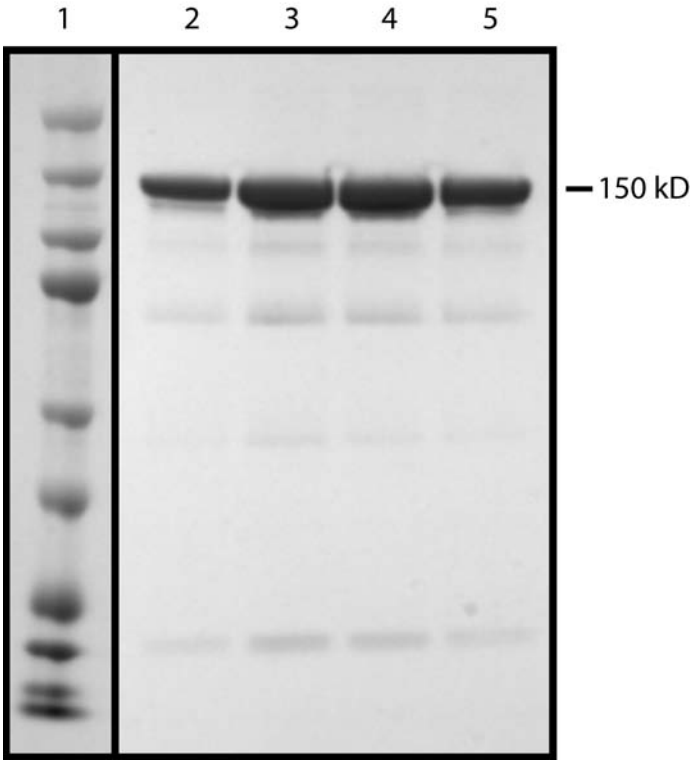


Figure 7. Protein gel electrophoresis of purified A20 monoclonal antibody. Protein was run against Kaleidoscope Precision Plus Protein Standard from Bio-Rad (lane 1). Protein G elution fractions are shown from 2-4 mL (lane 2), 4-6 mL (lane 3), 6-8 mL (lane 4), and 8-10 mL (lane 5).

ASSESSMENT OF A20 MAB FUNCTIONALITY

Rationale

The preparation of A20 monoclonal antibody was assessed to be the correct size according to protein gel electrophoresis. It was further necessary to characterize the MAb as being functionally relevant and capable of binding to AAV-2. The binding of monoclonal antibody A20 to AAV-2 is characterized here through a dot-blot assay.

Methods

Dot blots were used to assess A20 MAb biological functionality. AAV-2 was immobilized on a nitrocellulose membrane, in a second case was diluted into Triton-X before immobilization, and in a third case diluted into Triton-X and heated before immobilization. In all cases, the immobilized virus was blocked with 5% BSA in tris-buffered saline and Tween-20 (TBS-T). Purified A20 MAb was diluted into 0.1% BSA in TBS-T to a concentration of 1μg/mL and then incubated with the immobilized virus for 30 min at room temperature. The sample was washed and incubated with rabbit anti-mouse secondary antibody (Fc fragment specific) conjugated to alkaline phosphatase (Sigma-Aldrich). The sample was washed and then detected with 1-Step NBT/BCIP (Thermo Scientific) detection reagent until desired intensity was reached.

Additionally, 2.5 mg of A20 was biotinylated with the Amersham ECL Protein Biotinylation Module (GE Healthcare) according to the manufacturer's protocols. The dot-blot was then repeated in a second case where biotinylated A20 was used in place of primary and secondary antibodies and the sample was incubated with a 1:200 dilution of alkaline phosphatase conjugated streptavidin (Pierce) prior to incubation with detection reagent.

Results

It has been reported that A20 binds to a conformational epitope on the surface of the AAV-2 capsid and does not bind to individual VP-3 capsid proteins (Wobus et al.,

2000). Dot-blot analysis showed that A20 bound strongly to immobilized AAV-2 (Figure 8). As expected, A20 also bound strongly to AAV-2 in the presence of Triton-X detergent, where AAV-2 is still present in its native form. However, when AAV-2 was heated in the presence of Triton-X the capsid was denatured, and hence A20 did not bind strongly. This is consistent with Wobus et al.'s previous report.

Biotinylation was used as a method of detection for A20 in several later experiments. Dot-blot analysis of AAV-2 using biotinylated A20 yielded a similar result (Figure 9). Therefore, biotinylation does not alter the biological functionality of the antibody from the native state.

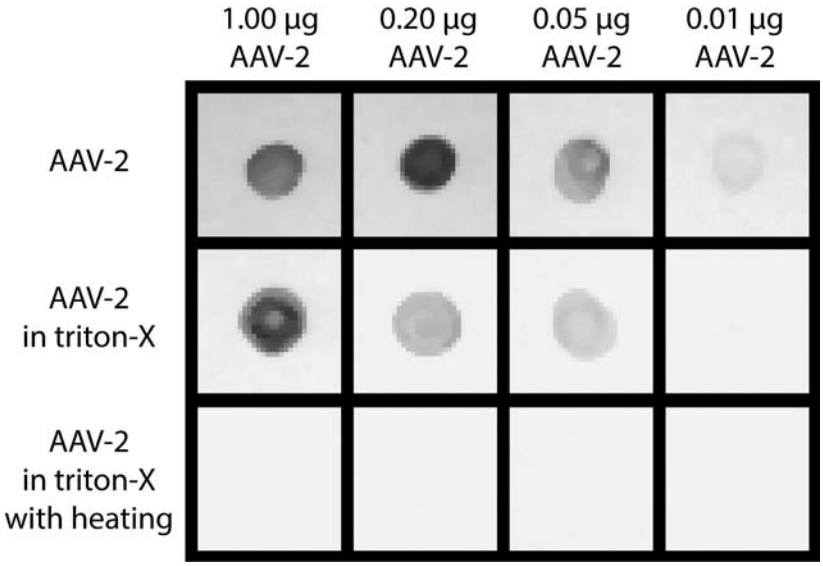


Figure 8. Dot-blot of immobilized AAV-2 incubated with A20 MAb primary antibody and AP-conjugated secondary antibody. Native AAV-2 is shown in the top row in decreasing concentrations (left to right). Native AAV-2 with Triton-X present is shown in the middle row in decreasing concentrations (left to right). Denatured AAV-2 with Triton-X present is shown in the bottom row in decreasing concentrations (left to right).

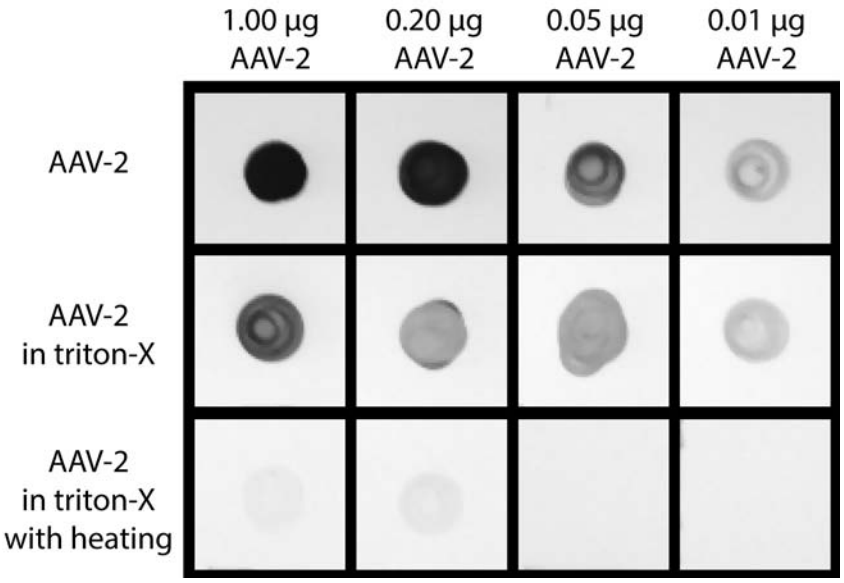


Figure 9. Dot-blot of immobilized AAV-2 incubated with biotinylated A20 MAb. Native AAV-2 is shown in the top row in decreasing concentrations (left to right). Native AAV-2 with Triton-X present is shown in the middle row in decreasing concentrations (left to right). Denatured AAV-2 with Triton-X present is shown in the bottom row in decreasing concentrations (left to right).

PRODUCTION AND PURIFICATION OF A20 FAB

Rationale

As discussed prior, the structural determination of antibody-virus complexes are usually accomplished by first cleaving the Fab domains from the monoclonal antibody (Hafenstein et al., 2009; Kaufmann et al., 2007; Smith et al., 1993). The Fab domain is monovalent (as opposed to the bivalent MAb) which is necessary for structural determination through single-particle *cyro*-EM. The production of A20 Fab and initial assessment of the product through protein gel electrophoresis follows here.

Methods

A20 Fab was prepared by digesting the antibody with papain. 10 mg of purified antibody was incubated with 250 µL of immobilized papain cross-linked to beaded agarose (Thermo Scientific) for 5 hrs at 37 °C. The immobilized papain was then removed from the solution by centrifugation at 1,000 RCF for 1 minute. Remaining functional A20 monoclonal antibody was removed from the solution by passing over a protein A affinity column with 20 mM sodium phosphate running buffer at pH 7.0. The column was cleaned by eluting bound MAb and Fc with 100 mM citric acid.

Results

The standard digestion protocol of antibody with papain was designed by Pierce. However, it was determined that the recommended protocol unexpectedly degraded the Fab. The protocol includes the use of cysteine during digestion which immediately starts dissociation of the A20 antibody (Figure 10). Buffer controls show no degredation. Cysteine (alone) degrades A20 into bands at 150kD, 125kD, 75kD, 50kD, and 25 kD which is consistent with interchain disulfide bond reduction. Reduction of A20 was apparent at cysteine concentrations as low as 0.8 mM, which is 25-fold lower than recommended by the Pierce protocol. The protocol was adjusted by removing cysteine from the digestion.

Protein gel electrophoresis of the revised papain digest showed strong doublet bands at 50 kD (Figure 11), which is the expected molecular weight for an Fab. Doublet bands were not expected from the protein digestion and purification. The

Fc region of the antibody also has an apparent molecular weight of 50 kD, however, functional Fc was being removed by Protein A purification. It is possible that the Fab is in two forms, where one contains a reduced intrachain disulfide and the other is in the native form, which may explain the apparent molecular weight difference exhibited from protein gel electrophoresis. However, it is surprising that both the upper and lower bands appear to have equivalent intensities, even when replicate experiments are performed.

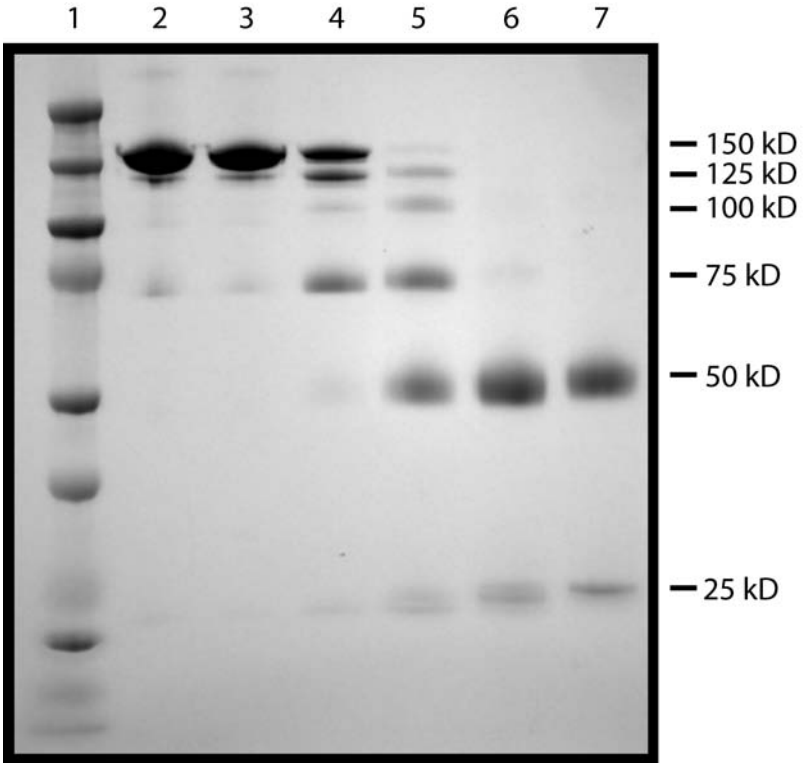


Figure 10. Protein gel electrophoresis of A20 monoclonal antibody incubated with varying amounts of cysteine. Protein was run against Kaleidoscope Precision Plus Protein Standard from Bio-Rad (lane 1). Samples are shown of A20 in water (lane 2), 20 mM PBS and 10 mM EDTA (lane 3), 0.8 mM cysteine (lane 4), 5 mM cysteine, (lane 5), 20 mM cysteine (lane 6), and 80 mM cysteine (lane 7).

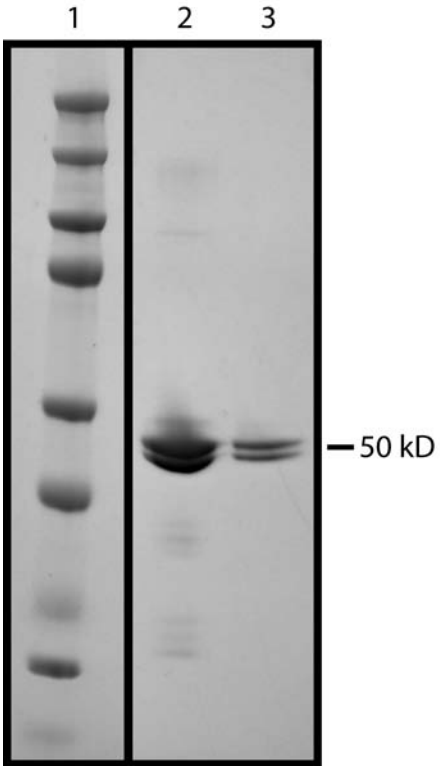


Figure 11. Protein gel electrophoresis of purified A20 monoclonal antibody digested with papain. Protein was run against Kaleidoscope Precision Plus Protein Standard from Bio-Rad (lane 1). Purified digest was overloaded to show purity (lane 2) and diluted to show the discreet doublet band (lane 3).

MASS SPECTROMETRY OF A20 FAB

Rationale

Protein gel electrophoresis of the A20 Fab showed a strong doublet band, which is not a conventional result for MAb digestion. This gave rise to several questions including whether the doublet band was Fab or whether the upper and lower band contained the same protein. Therefore, mass spectrometry on the doublet band was used to characterize the bands according to their sequence. At the time of

experimentation the sequence of A20 was not known, thus results needed to be compared to sequences deposited in protein sequence data banks. To this effect, mass spectrometry was used to characterize both the upper and lower band of the A20 digest.

Methods

Gel electrophoresis was performed on purified A20 Fab using a 4% to 12% Bis-Tris gradient gel. Bands with a molecular weight of 45-50 kD were excised and sent to the OHSU Proteomics Core where the bands were digested and the resulting small peptide fragments were sequenced.

Results

Both bands were compared against known protein sequences in SWISS-PROT and TrEMBLE by peptide mass fingerprinting. There was no appreciable difference between the upper and lower band protein sequence matches. The sequence of the light chain of a blocking antibody to human Mcp-1 was the nearest protein match and showed 59-63 identified spectra matching A20 Fab (Table 2). Other identified proteins contained 10-11 or less identified spectra matching A20 Fab. The number of identified spectra was based on a 20% probability that the matching protein identification is correct as calculated by Scaffold 3 (Proteome Software). This suggests that both the upper and lower bands are identical in sequence and verifies that the purified protein is an IgG. It is possible that the doublet band contains the

Fab in two different forms, one with a reduced intra-chain disulfide bond and one native, where the reduced form would exhibit an increased apparent molecular weight in protein gel electrophoresis.

	Sequence
Upper Band	DIQMTQSSSSFSVSLGDRVTITCK----------LLISGATSL ETGVPSRFSGSGSGK----------R ADAAPTVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSERQNGVLNS WTDQDSKDSTYSMSSTLTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRNEC
Lower Band	DIQMTQSSSSFSVSLGDRVTITCK----------LLISGATSL ETGVPSRFSGSGSGK----------R ADAAPTVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVK--------QNGVLDS WTDQDSKDSTYSMSSTLTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRNEC

Table 2. Sequence the light chain of a blocking antibody to human Mcp-1. Amino acids showing over 80% probability of matching the A20 Fab spectra are listed with their sequence. Amino acids with less than 80% probability of matching the A20 Fab sequence are marked with dashes.

ASSESSMENT OF A20 FAB FUNCTIONALITY

Rationale

The determination that prepared A20 Fab sequence matched homologous Fab sequences by peptide mass fingerprinting was not sufficient to ensure that the Fab was functionally relevant. There was the possibility that the Fab was denatured or in another functionally irrelevant form where it would not bind to AAV-2. Four independent methods were used to assess the functionality of A20 Fab relative to AAV-2.

Competition Dot-Blot Assay

Rationale

The first method of assessing the functionality of AAV-2 was to use a competition dot-blot assay. Here A20 Fab was used in an attempt to compete off A20 monoclonal antibody. While this technique was not successful it confirming the functionality of A20 Fab, it should be noted that this technique is later used successfully in confirming the functionality of A20 Fab'.

Methods

To assess whether purified A20 Fab was functional, a competition dot-blot assay was performed. AAV-2 was incubated overnight with varying quantities of Fab at 4°C. The Fab-AAV-2 mixture was then immobilized on a nitrocellulose membrane. The immobilized mixture on the membrane was blocked with 5% BSA in TBS-T. Purified A20 MAb was diluted into 0.1% BSA in TBS-T to a concentration of 1 μg/mL and then incubated with the immobilized virus for 30 min at room temperature. The sample was washed and incubated with rabbit anti-mouse secondary antibody (Fc fragment specific) conjugated to alkaline phosphatase (Sigma-Aldrich). After washing, the sample was then detected with 1-Step NBT/BCIP (Thermo Scientific) detection reagent until desired intensity was reached.

Results

Various amounts of A20 Fab were pre-incubated with 0.2μg of AAV-2 before immobilization in an attempt to block exposed A20 epitopes. Immobilized virus and varying amounts of A20 Fab were incubated with 10 ng/μL of A20 MAb, subsequently with a 5000-fold dilution of anti-mouse Fc specific secondary antibody (Sigma-Aldrich), and then detected with 1-Step NBT/BCIP (Thermo Scientific) detection reagent until desired intensity was reached. In the case that A20 epitopes were blocked by bound A20 Fab (indicating formation of Fab-AAV-2 complex), these complexes would exhibit reduced signal when incubated with A20 primary antibody.

Experimental results indicated there was no reduction in primary antibody signal, regardless of the amount of Fab that AAV-2 was pre-incubated with (Figure 12). Note that when this assay was repeated with Fab', a reduction in intensity was seen corresponding to higher Fab' concentrations (*vide infra*). This suggests two possible scenarios are occurring: (1) Purified A20 Fab not does binding strongly to AAV-2 before immobilization on the nitrocellulose membrane or (2) A20 primary antibody is competing off bound A20 Fab during the related incubation step. It is not experimentally feasible to experimentally distinguish which of the two scenarios is occurring, although scenario (1) is consistent through all Fab functionality assessment trials (*vide infra*). However, scenario (2) remains a significant possibility, due to the 400-fold lower binding affinity of A20 Fab' as compared to A20 MAb.

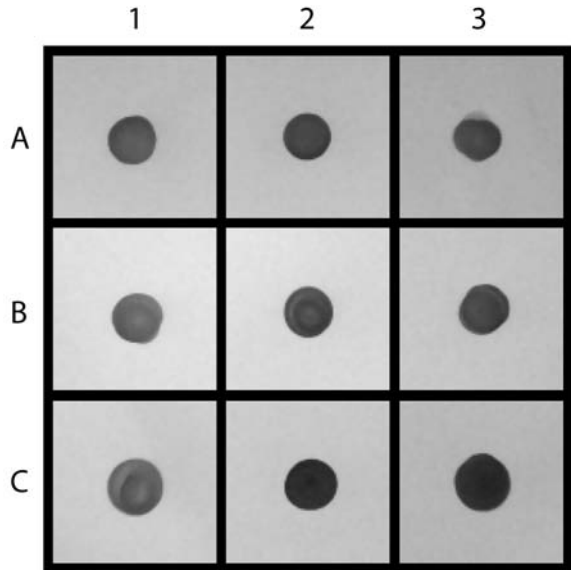


Figure 12. Dot-blot of AAV-2 and varying quantities of A20 Fab immobilized on a nitrocellulose membrane and incubated with A20 primary antibody and an Fc specific secondary antibody. AAV-2 was immobilized on nitrocellulose without A20 Fab (A1) and pre-incubated with a 10:1 stoichiometric ratio of A20 Fab to AAV-2 (A2), a 20:1 ratio (A3), a 30:1 ratio (B1), a 40:1 ratio (B2), a 50:1 ratio (B3), a 60:1 ratio (C1), a 240:1 ratio (C2), and a 600:1 ratio (C3).

Biotinylated Dot-Blot Assay

Rationale

Precise interpretation of negative results for the previous competition assay can be difficult. Therefore, an alternative means to verify that A20 Fab was capable of binding to AAV-2 was sought, ideally one which directly detects the Fab. Dot-blot assays of AAV-2 showed excellent signal when using biotinylated A20 MAb as a primary antibody, so biotinylated A20 Fab was considered a possible means of

detecting the A20 Fab as well. Accordingly, the functionality of A20 Fab is assessed here via biotinylation on a dot-blot assay.

Methods

A20 Fab was dialyzed into PBS and then 2.5 mg was biotinylated with the Amersham ECL Protein Biotinylation Module (GE Healthcare) according to the manufacturer's protocols. A dot-blot assay was performed as described in "Assessment of A20 MAb Functionality", with the exception that the biotinylated A20 Fab replacing A20 MAb obviated the need for secondary antibody.

Results

The dot-blot assays showed no significant signal (Figure 13). In a minimal positive control, dot-blot assays of immobilized, biotinylated A20 Fab showed excellent signal when incubated with detection reagent. Several possible scenarios may contribute to the lack of signal: (1) Biotinylation interfered with Fab binding or (2) the Fab is not capable of binding to AAV-2. It is not feasible to distinguish which of these two scenarios is occurring. However, given that failure of Fab to bind AAV-2 would explain both these results and the MAb competition assay, this seems the more likely interpretation.

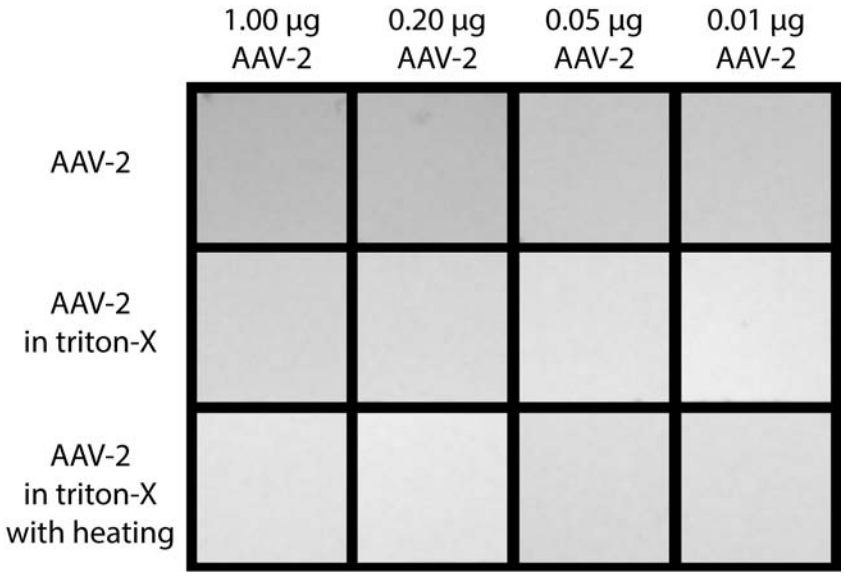


Figure 13. Dot-blot of immobilized AAV-2 incubated with A20 Fab that has been through the biotinylation process. Native AAV-2 is shown in the top row in decreasing concentrations (left to right). Native AAV-2 with Triton-X present is shown in the middle row in decreasing concentrations (left to right). Denatured AAV-2 with Triton-X present is shown in the bottom row in decreasing concentrations (left to right).

Size-Exclusion Chromatography

Rationale

Given the central importance of preparing monovalent antibody fragments and increasing doubts about the functionality of the prepared Fabs, size exclusion chromatography was attempted as a third assay. The idea was that unbound Fab and AAV-2 would have very different elution times, but if Fab were bound (even transiently), then at least a fraction of the Fab would have changed elution as it co-migrated (at least temporarily) with the virus.

Methods

AAV-2 was dialyzed into PBS and incubated overnight with A20 Fab at 4°C. The solution was then passed over a size exclusion column of Sephacryl S-200 (GE Healthcare) with a bed volume of 24 mL. The running buffer contained 5% HMG and the flowthrough was collected in 2 mL fractions.

Results

A column containing Sephacryl S-200 has resolving capability between 5 kD and 250 kD. AAV-2 has a molecular weight of approximately 4,000 kD and A20 Fab is approximately 50 kD. Hence, it would be expected that both AAV-2 and A20 bound to AAV-2 would exit in the void volume of the column and unbound Fab would elute in the latter portion of the column's resolving range. In the case of irreversible binding, bands of AAV-2 and A20 would be present in relative amounts equivalent to the stoichiometric binding of A20 where complex formation was occurring.

It was observed that there was a significant overlap between the elution of the A20 and AAV-2 and there was a difference in the flow-through average volume in which the two exited (Figure 14). In the absence of controls discussed below, only a preliminary interpretation can be made. It is not known whether the virus alters the Fab elution, However, the initial flowthrough of virus does not contain Fab, raising the concern that binding may not be stable enough for cryo-EM. Lack of observed binding via size exclusion chromatography may be due to several scenarios: (1) Previously purified Fab does not bind strongly to AAV-2, (2) dynamic equilibrium

(3) there was nonspecific binding of the Fab to sephacryl beads used for the size exclusion chromatography. Scenario (1) remains a consistent possibility through all Fab functionality assessment trials. However, scenario (2) and (3) are significant possibilities and are compounded by the 400-fold reduced affinity of the Fab', which makes both the dissociation and non-specific binding of the Fab considerably more likely.

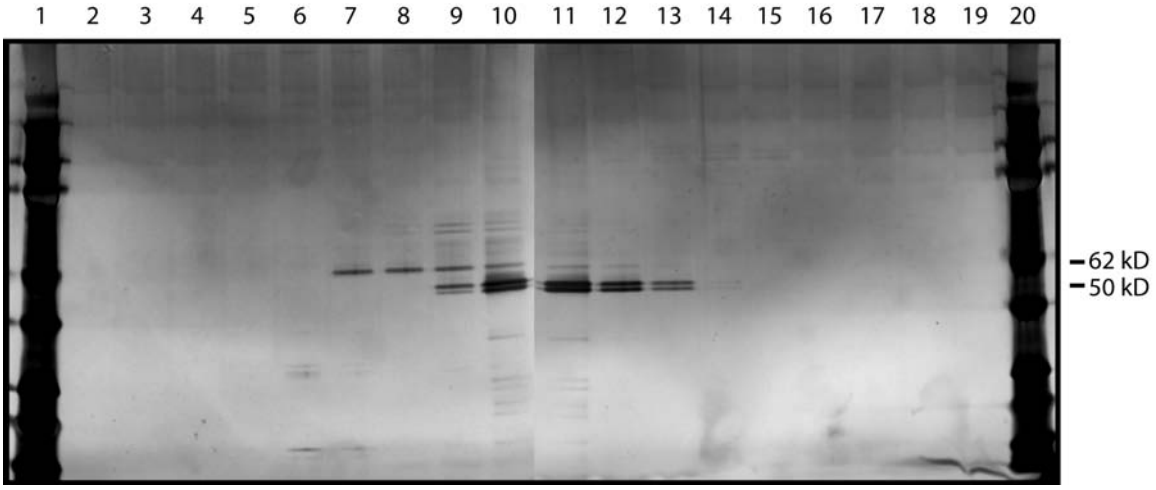


Figure 14. Size-exclusion chromatography of AAV-2 pre-incubated with A20 Fab. Protein was run against Kaleidoscope Precision Plus Protein Standard from Bio-Rad (lane 1 and 20). Fractions of 0.5 mL chronologically progress from left to right (lane 2 – 19).

Sucrose Cushion

Rationale

Size-exclusion chromatography of A20 Fab and AAV-2 yielded results suggesting that A20 Fab did not bind strongly to AAV-2. As a second, independent verification

experiment, a sucrose cushion was also used to detect A20 Fab bound to AAV-2. Unlike size-exclusion chromatography, the sucrose cushion was also a viable means to isolate Fab-AAV-2 complex at the concentrations needed for *cyro*-EM.

The principles would be similar to size exclusion chromatography: detection and isolation of complexes where Fab has co-migrated with AAV, separated in this case by the difference in sedimentation velocity between small isolated Fab, and large viral complexes.

Methods

50 μg of AAV-2 was incubated with a 4-fold stoichiometric excess of Fab overnight at 4 °C. 1 mL of 30% sucrose was layered on top of 1 mL of 60% sucrose to create a sucrose cushion. The Fab/AAV-2 sample was layered on top of the sucrose cushion and centrifuged for 1 hr at 32,000 RCF. The bottom of the centrifugation tube was punctured and 0.5 mL fractions were collected by gravity elution.

Results

Fab and AAV-2 were incubated overnight prior to its purification in the sucrose cushion. In a sucrose cushion a larger mass to surface area ratio will contribute to faster migration down the sucrose gradient. Therefore, uncomplexed A20 Fab was expected to reside on the top of the first cushion after centrifugation, while AAV-2 and complexed Fab was expected to reside on the top of the second cushion.

Analysis by protein gel electrophoresis show that AAV-2 migrated faster than Fab (Figure 15). It is unclear if a stable Fab-AAV-2 complex formed. Note that there is evidence that Fab migration slowed, suggesting formation of at least a transient complex. However, most of the AAV comes through in an uncomplexed form (lane 6). As there is apparently excess Fab (lane 10), there are several possibilities that may contribute to an unstable complex including (1) dynamic equilibrium, (2) populations of binding and disordered Fab, or (3) it is nearly impossible to saturate the 60 epitopes of AAV-2. Regardless of the reason for instability, the Fab-AAV-2 complex is ill-suited for structural assessment via cyro-EM.

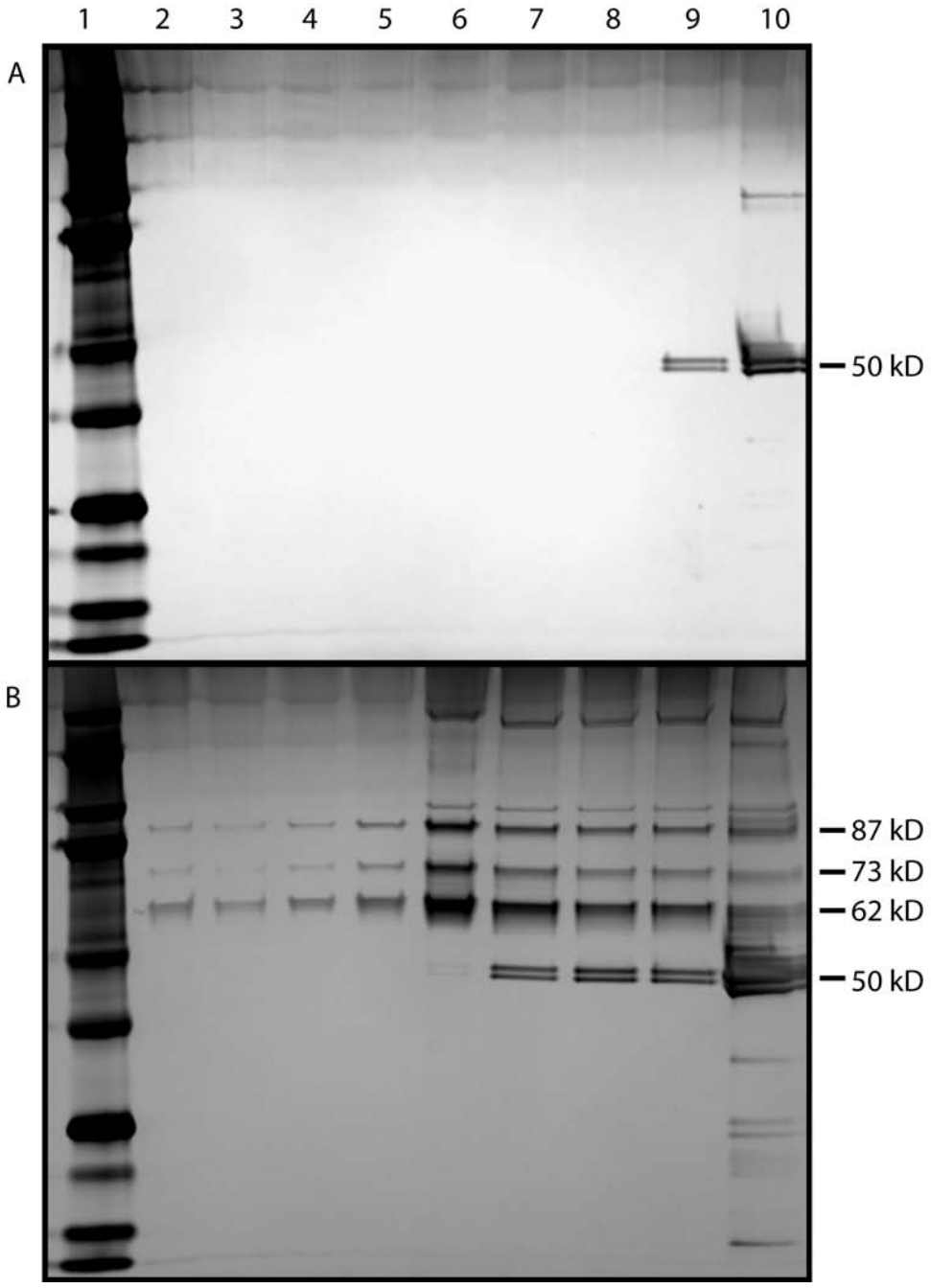


Figure 15: Sucrose cushion of A20 pre-incubated with AAV-2. Protein was run against Kaleidoscope Precision Plus Protein Standard from Bio-Rad (lane 1). The control sucrose cushion contains A20 Fab and no AAV-2 (pane A) and the experimental sucrose cushion contains A20 Fab pre-incubated with AAV-2 (pane B). Lowermost cushion fractions are present on left (lane 2), the 30%-60% sucrose

cushion interface is present in the middle (lane 6), and progresses to the uppermost cushion fractions on the right (lane 10).

ELECTRON MICROSCOPY OF AAV-2 COMPLEXED WITH A20 MAB

In spite of several threads that were not completed with the manipulation and assessment of A20 Fab, priorities were shifted to other approaches that appeared to have more promise. Two approaches were attempted, one of which was using *cryo*-EM to image AAV-2 with whole, undigested A20 antibody, which will be addressed in this section. In the second approach, the preparation of alternative cleavage products for single-particle *cryo*-EM was attempted, which will be addressed in "Structure of Adeno-associated Virus-2 in Complex with Neutralizing Monoclonal Antibody A20".

Rationale

Results from functional characterization of A20 Fab suggested that digesting the monoclonal antibody with papain reduced the binding affinity to non-ideal levels. In order to maintain functionality of the antibody, attention was focused on imaging native monoclonal A20 antibody in complex with AAV-2. It was our initial objective to achieve a moderate-resolution reconstruction of native monoclonal A20 antibody bound to AAV-2, although an alternate approach was eventually adopted for reconstruction.

Imaging of complexes under single-particle conditions requires that each particle or complex be isolated from other particles. Unfortunately, when mobile AAV-2 and mobile A20 MAb are present in the same solution, a precipitate immediately forms of cross-linked viral particles. The SMART Grid technology by Dune Sciences, Inc. (unpublished) offered the prospect of immobilizing either AAV or the antibody on the surface of an EM grid, where the binding partner could then be added without fear of aggregation. SMART Grids are comprised of a silicon wafer with silicon dioxide windows that have a functional group bound.

SMART Grids functionalized with protein A and lysine were of interest to this research. In the first case, it is possible to use protein A to capture A20 antibody, which can then capture AAV-2. In the second case, it is possible to cross-link AAV-2 to lysine with a crosslinker such as bis[sulfosuccinimidyl] suberate. Both of these methodologies serve to immobilize AAV-2 to a surface that may be directly imaged through electron microscopy. Once AAV-2 is immobilized, it may be treated with A20, secondary antibodies, and washing without loss of product. Thus, SMART Grids allow the circumvention of conventional issues that arise from imaging native antibody with antigen.

Methods

AAV-2 Complexed with A20 MAb on Protein A SMART Grids

Preliminary experiments started with a test of on-grid complex formation by sandwiching AAV-2 between immobilized MAb A20 and freely bound MAb A20, detecting binding of the latter by immunogold labeling. First, SMART Grids functionalized with Protein A were incubated with 0.5 μg/mL of A20 for 5 min. Excess A20 was removed by washing. Grids were then incubated in 0.01 mg/mL of AAV-2 in 5% HMG for 5 min where bound A20 served to immobilize AAV-2 on the grid. Excess AAV-2 was removed by washing. Grids were incubated in 0.01 mg/mL of A20 MAb in 5% BSA for 5 min where the monoclonal A20 bound to the immobilized AAV-2. Excess A20 MAb was removed by washing. Grids were then incubated with a 10-fold dilution of anti-mouse IgG labeled with 5 nm gold. Excess secondary antibody was removed by washing.

In addition, a control was performed that was identical to the above protocols, with the variation that 5% HMG was used in place of 0.01 mg/mL of AAV-2 in 5% HMG for the viral incubation step.

AAV-2 Complexed with A20 MAb on Lysine SMART Grids

A second approach was taken to visualize where AAV-2 was immobilized on lysine SMART Grids. First, SMART Grids functionalized with lysine were prepared for cross-linking by incubating the grid in 10 mg/mL of bis[sulfosuccinimidyl] suberate

(BS3) for 45 min. Excess BS3 was removed by washing. Grids were then incubated in 0.01 mg/mL of AAV-2 in 5% HMG for 30 min where bound BS3 served to cross-link AAV-2 to the grid. Excess AAV-2 was removed by washing. Grids were incubated in 0.01 mg/mL of A20 MAb for 30 min where the monoclonal A20 bound to the immobilized AAV-2. Excess A20 MAb was removed by washing. Grids were then incubated with a 10-fold dilution of anti-mouse IgG labeled with 5 nm gold (Sigma-Aldrich). Excess secondary antibody was removed by washing.

Results

AAV-2 Complexed with A20 MAb on Protein A SMART Grids

SMART Grids from Dune Sciences were used to image AAV-2 in the presence of bivalent monoclonal A20 antibody. SMART grids functionalized with Protein A provided a foundation for A20 MAb to be bound to the grids, which subsequently captured AAV-2. Low concentrations A20 were used to capture AAV-2, in order to prevent false-positives in future labeling steps. Immobilized AAV-2 was labeled with A20 MAb and gold-labeled secondary antibody specific to IgG. Controls that excluded immobilized AAV-2 particles exhibited a relatively few nanogold particles (Figure 16). Additionally, images of AAV-2 incubated with A20 antibody and secondary antibody show a significantly higher concentration of nanogold present in the immediate vicinity of AAV-2 particles than in the gulfs between viral particles. This suggests that A20 monoclonal antibody forms a complex with AAV-2 on SMART Grids.

AAV-2 Complexed with A20 MAb on Lysine SMART Grids

Complexes of AAV-2 and A20 MAb were created on lysine SMART Grids in a manner very similar to that which was used for protein A SMART Grids. The primary variation involved using the cross-linking reagent BS3 to immobilize AAV-2 instead of using a capture antibody. The resultant images for lysine grids had a greater variation in quality than the protein A SMART grid analyses. However, resultant images led to instructive qualitative results. Primarily, images were acquired that distinctively showed up to 6 or 8 bound nanogold particles per virus (Figure 16). Depending on the occupancy of the secondary antibody, the actual number of bound A20 monoclonal antibodies may have been higher.

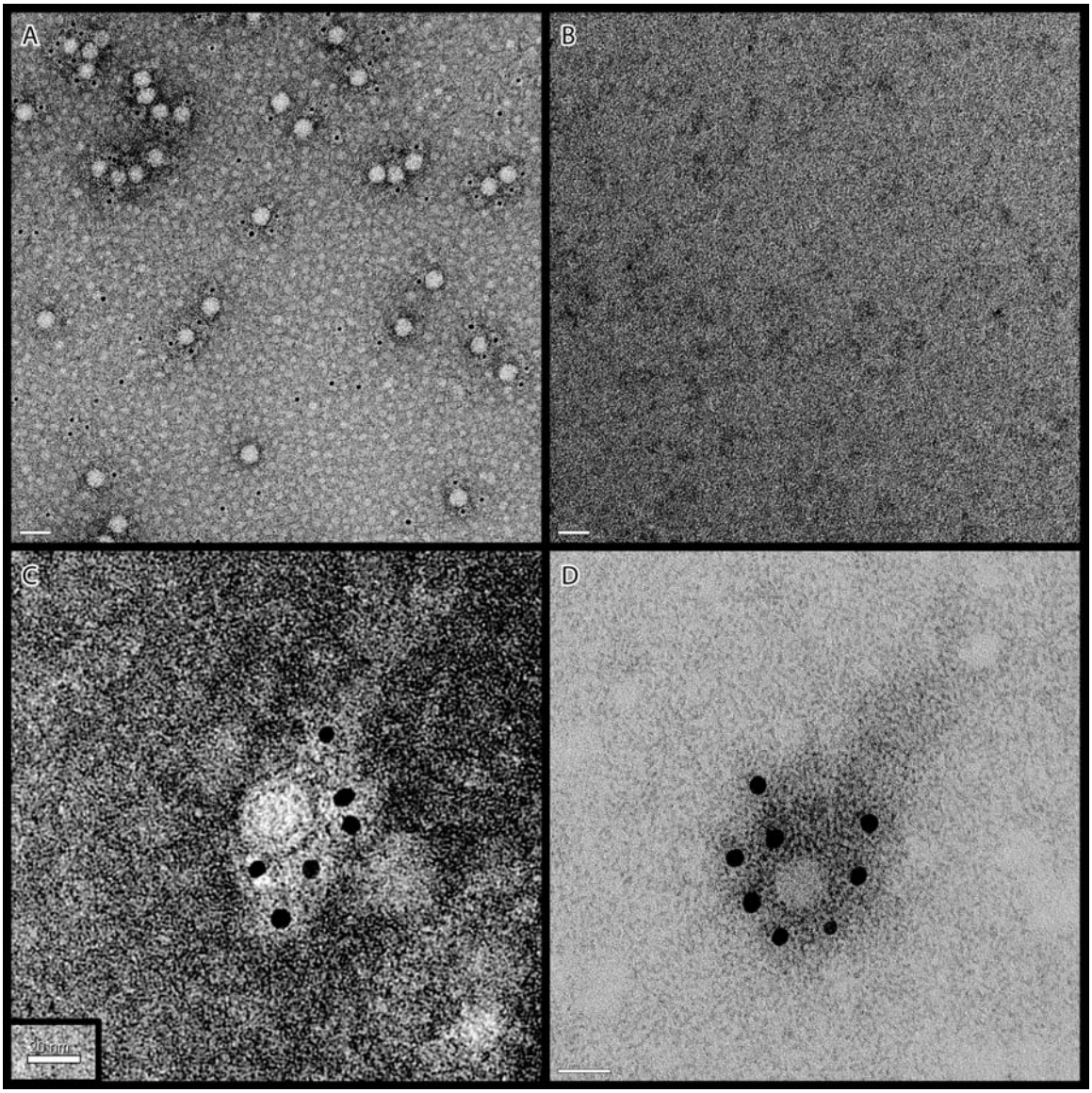


Figure 16. Negative Stain Electron Microscopy of AAV-2 complexed with A20 MAb. Images of AAV-2 bound to protein A SMART Grids and incubated with A20 primary MAb (mouse) and mouse-specific, nanogold-labeled secondary antibody show nanogold particles in the immediate vicinity of AAV-2 (panel A). Images of protein A SMART Grids incubated with A20 MAb (mouse) and mouse-specific, nanogold-labeled secondary antibody show a relatively low background of 2 nanogold particles when no AAV-2 is present (panel B). AAV-2 was cross-linked to lysine SMART Grids and incubated with A20 MAb and nanogold-labeled secondary antibody, where up to 6 or 8 nanogold particles were visible per virus (panel C and D).

DISCUSSION

Preparation of AAV-2 and MAb A20

The production and purification of adeno-associated virus 2 and A20 monoclonal antibody were successful, with each yielding functional products. Several new facts were learned about A20 monoclonal antibody. First, A20 self-associates and precipitates at high concentrations when exposed to temperatures of approximately 4 °C, similar to that of cryoglobulins, but does not self-associate at 25 °C. Second, A20 is reduced in the presence of cysteine. The concentration of cysteine that may be present without reducing A20 antibodies is at least 25-fold lower than standard antibody digestion protocols.

Preparation of Fab fragments

The production and purification of A20 Fab was only superficially successful, with the resultant product not in a viable form for use in the creation of a stable Fab-AAV-2 complex. Upon reduction of the A20 monoclonal antibody, a doublet band is formed of approximately 50 kD, which is the expected size of immunoglobulin Fabs. Mass-spectrometry results from peptide mass fingerprinting suggest that the doublet band contained a single protein sequence. The matching sequences are from IgG sequences in the protein data bank.

Assessment of the functionality of purified A20 Fab proved difficult. Four methods employed showed the absence of binding or the presence weak binding, at most.

Four points can be drawn about the Fab from these experiments: (1) AAV-2 pre-incubated with A20 Fab does not inhibit binding of monoclonal A20. (2) Biotinylated A20 Fab does not show significant binding to AAV-2. (3) AAV-2 pre-incubated with A20 Fab does not elute with early AAV-2 fractions when purified via size-exclusion chromatography or via a sucrose cushion, signifying that A20 Fab does not bind strongly to AAV-2. (4) Results from attempted isolation of Fab-AAV-2 complex via a sucrose cushion suggest that Fab weakly binds to AAV-2, although it is not known whether binding is specific or non-specific. Although caution should be taken in the interpretation of fundamentally negative results, and although analysis by size exclusion chromatography was preliminary, without full controls, the weight of evidence was that Fab would likely not be a suitable reagent for the production of stable complex for cryo-EM visualization, and that other avenues should be investigated.

Electron Microscopy of AAV-2 Complexed with A20 MAb

To the best of our knowledge, this is the first time that a complex of native monoclonal antibody and virus has been imaged with functionalized electron microscopy grids. Imaging negatively stained AAV-2 with bound nanogold particles was a preliminary step taken to elucidate the feasibility of higher-resolution structural studies. Additionally, this work was performed with Dune Sciences and John M. Miller to establish proof-of-concept of a robust and expedient methodology for viral imaging with SMART Grid technology. This work was successful as a proof

of concept and the methodology is currently advertised on their website (as of 25 April 2012): *http://www.dunesciences.com/apps_aav2.php*.

With regard to moderate-resolution imaging, complexes of AAV-2 and A20 on functionalized EM grids can theoretically be reconstructed into a 3-D electron density map. However, the current technology, as presented here, is not expected to be sufficient for a moderate-resolution reconstruction of the complex. In order for reconstructions to be performed on samples at higher resolution, we expect that three advancements must be pursued with the current technology.

First, the thickness of the functionalized grids must be decreased. The relatively thick nature of functionalized grids increases the background noise to unfavorable levels, which will hinder many aspects of the reconstruction process. In 2011, and after the conclusion of this study, Dune Sciences appears to be addressing this concern with the recent release of a thinner, functionalized grid called C-SMART grids.

Second, the grid strength must be strong enough to reliably withstand freeze-plunging. Preliminary experiments showed that the silicon dioxide windows broke during freeze-plunging, which is a necessary for moderate-resolution studies under cryogenic conditions.

Third, a methodology must be developed to create a local environment where antigenic epitopes are near occupancy limits and relatively low background is maintained. Currently, observed occupancy is up to 6-8 monoclonal antibodies per

virus for lysine SMART Grids, although actual occupancy may be higher. Steric constraints will limit the occupancy, which include steric interference from the grid itself and additional interference from monoclonal antibodies that are 3-fold larger than their Fab counterparts.

Protein A grids currently have the benefit of relatively homogenous imaging results and use a capture mechanism which is also the binding mechanism under study. Unfortunately, protein A grids do not currently appear to have the optimal occupancy that is present for the lysine grids. However, sample preparation was constrained by use of labeled-secondary antibody which detects A20 bound directly to the SMART Grid. This limited preparation techniques to those that kept unbound A20 to minimal levels. In the case where a labeled secondary antibody was not necessary, conditions could be optimized for higher occupancy.

Perhaps most intriguing future prospect of this technology lies in its potential designer qualities. Most ideal would be functionalizing a specific protein of choice to an EM grid. Methods to produce individually specialized grids on a thin surface may allow the rapid imaging of complexes, such as AAV-2 to receptors and antibodies, near moderate-resolution conditions.

Overall, preliminary results using SMART Grid technology were encouraging. A20 MAb was shown to bind to AAV-2 in an environment suitable for electron microscopy imaging. SMART Grid technology is currently suitable for preliminary antigen-virus studies and will likely yield moderate-resolution structural results with the proper improvements in the technology. Progress was also made along

other avenues, however, which gave promise for more expedient and higher resolution results with regards to a moderate-resolution structure of A20 bound to AAV-2.

STRUCTURE OF ADENO-ASSOCIATED VIRUS-2 IN COMPLEX WITH NEUTRALIZING MONOCLONAL ANTIBODY A20

CONTRIBUTIONS

Dustin M. McCraw, Jason K. O'Donnell, Kenneth A. Taylor, Scott M. Stagg, and Michael S. Chapman contributed to this manuscript. Dustin McCraw performed sample preparation, particle picking, and analysis of the completed reconstruction. Dustin McCraw and Jason O'Donnell jointly performed grid preparation and data collection. Jason O'Donnel performed reconstruction of the complex.

INTRODUCTION

Gene therapy involves the delivery to cells of DNA designed to alleviate disease of genetic nature or predisposition, DNA that commonly encodes a functional replacement for a mutated gene. AAV has become a leading candidate vector for gene therapy because it is non-pathogenic, induces little inflammatory response, integrates site-specifically into chromosome 19, infects dividing and non-dividing cells, has wide cell tropism, and is producible in large quantity (Carter, Burstein, and Peluso, 2008). AAV's main disadvantages include a limited gene payload of 5kb (Dong, Fan, and Frizzell, 1996) and seropositivity of 35-80% of the U.S. population to AAV-2 (Calcedo et al., 2009) (*vide infra*). Current and recent clinical trials using AAV vectors are targeting hemophilia B, prostate and melanoma cancers, Canavan

disease, Alzheimer's, Parkinson's, muscular dystrophy, rheumatoid arthritis and HIV vaccines (Carter, 2005).

AAV is a non-enveloped virus with a single-stranded DNA genome surrounded by a protein capsid that is composed of 60 protein subunits related by icosahedral symmetry (Xie et al., 2002). Three capsid proteins; VP-1, VP-2, and VP-3 are present in a ratio of 1:1:10, and have molecular weights of 87kD, 73kD, and 62kD, differing (only) in N-terminal extensions of 137 residues (VP-1) and 64 residues (VP-2) relative to VP-3 (Agbandje-McKenna and Kleinschmidt, 2011). It is not clear whether VP-1 & 2 substitute for VP-3 at specific locations in the otherwise 60-fold symmetric capsid.

Immune responses to AAV are mild and non-inflammatory, significantly reducing the possibility of complications during treatment (Bueler, 1999; Chirmule et al., 1999). However, immune response still presents a significant barrier to efficient delivery of the vector during initial treatment, as neutralizing antibodies have been found in up to 60% of the population (Blacklow, Hoggan, and Rowe, 1968; Boutin et al., 2010; Calcedo et al., 2009; Chirmule et al., 1999; Moskalenko et al., 2000). Clinical trials involving experimental cystic fibrosis and hemophilia treatments required re-administration of therapies (Flotte and Carter, 1998; Kay et al., 2000). As only ~10% of progenitor cells below the lung epithelium are transduced, transgene expression falls over 6 months as surface cells are turned over (Carter and Flotte, 1996). Exposure to natural AAV or vectors in animals and humans results in development of neutralizing antibodies (Halbert et al., 2000; Moskalenko

et al., 2000; Peden et al., 2004; Xiao et al., 1999). This can decrease the efficiency of transgene expression upon subsequent administration(s) of vectors of the same serotype (Davidoff et al., 2005; Glas and van der Linden, 2010; Manno et al., 2006; Wang et al., 2011). Cellular immune response is largely directed against the viral capsid, while immune response to the transgene product is rare (Brockstedt et al., 1999; Chirmule et al., 2000; Halbert et al., 1997; Hernandez et al., 1999; Manning et al., 1998). Therefore, it is widely thought that the engineering of antigenic variant vectors will have substantial impact in the development of efficient gene therapy treatments (Flotte, 2005; Peden et al., 2004; Xiao et al., 1997).

A20 is a monoclonal antibody which neutralizes AAV-2 and AAV-3B subsequent to primary receptor binding (Wobus et al., 2000). It does not bind to the other predominant serotypes 1, 4, 5, 6, 8 and 9 (Murphy et al., 2008; Wobus et al., 2000; Wu, Asokan, and Samulski, 2006; Xie et al., 2011). It is the most widely studied monoclonal antibody against AAV-2, due in part to its ability to bind only to fully assembled capsid and not to unassembled capsid proteins (Moskalenko et al., 2000; Wistuba et al., 1995; Wobus et al., 2000). It has been suggested that there are at least three neutralizing immunogenic sites on the capsid of AAV-2 (Lochrie et al., 2006), therefore it is unlikely that mutations solely within the A20 footprint will yield a capsid that is completely distinct antigenically. However, two mutations which inhibit A20 neutralization also confer resistance to polyclonal neutralization, suggesting that A20 binding might model neutralization at a dominant epitope (Lochrie et al., 2006). There have been multiple attempts to locate the A20 epitope using various techniques, including PEPSCAN, peptide competition, peptide

insertions, and both site-directed and scanning mutagenesis (Girod et al., 1999; Lochrie et al., 2006; Moskalenko et al., 2000; Shi, Arnold, and Bartlett, 2001; Wobus et al., 2000; Wu et al., 2000). Lochrie *et al*. (2006) have drawn attention to inconsistencies between some of the experimental epitope mappings, and between some of the proposed immunogenic sites and the subsequent AAV-2 atomic structure (Xie et al., 2002), noting that inaccuracies result from known limitations of standard mehods (Van Regenmortel, 1992). There has long been an interest in resolving these questions through direct visualization of an AAV-2/A20 complex (Lochrie et al., 2006).

We present here a *cryo*-EM reconstruction of AAV-2 in complex with A20 monoclonal Fab' antibody at a resolution of 8.5 Å. The antibody-binding footprint of AAV-2 has been determined from the previously determined crystal structure of AAV-2 by fitting a homology model of A20 Fab' into the experimental density of the complex. To our knowledge, this is the first structure of an AAV-antibody complex. The binding interface is compared with the corresponding surfaces of closely related viruses for insights into the characteristics of serotypes that might determine whether or not a virus is recognized by A20.

METHODS

Preparation of AAV2 and A20 Fab'

AAV2 was produced as described previously (Xie et al., 2004). Stocks of hybridoma cells for A20 were generously provided by Jürgen Kleinschmidt. They were grown by seeding $25x10^6$ cells in 15mL of 90% RPMI media (Sigma-Aldrich) and 10% FetalClone 1 serum (Hyclone) in a CELLine CL 1000 bioreactor (Integra). The bioreactor was maintained using 98% RPMI and 10% FetalClone 1 serum. Antibodies were harvested after 1 week. Cells were pelleted out by centrifugation at 70x g for 5 min and the supernatant was passed through a 0.45 μm filter. Secreted antibodies were purified with a HiTrap protein G affinity column (GE Healthcare).

The antibody solution was dialyzed into 20 mM sodium acetate (pH 4.5) and digested with immobilized pepsin (Thermo Scientific) as per the manufacturer's protocol to yield $F(ab_2)'$. The $F(ab_2)'$ solution was dialyzed overnight into 20 mM sodium phosphate (pH 7.2). Undigested IgG and Fc fragments were removed using a HiTrap protein A affinity column (GE Healthcare). Flow-through fractions containing $F(ab_2)'$ were dialyzed into 150mM PBS and 5 mM EDTA (pH 7.2) using a Float-A-Lyzer G2 (Spectra/Por) with a molecular weight 50 kD weight cutoff.

The $F(ab_2)'$ solution was reduced with Mercaptoethylamine-HCl (2-MEA; Thermo Scientific) according to the manufacturer's protocol. The resultant crude Fab' solution was immediately purified by size exclusion chromatography with Superdex 200 (GE Healthcare) using a running buffer of 150 mM sodium chloride, 50 mM

sodium phosphate, and 5 mM EDTA (pH 7.2). Gel electrophoresis with silver staining was used to assess purity.

Preparation and cryo-EM of AAV2-A20 Fab' complex

AAV-2 in 100 mM HEPES, 50 mM magnesium chloride, and 5% glycerol (pH 7.2) was incubated with a 4-fold excess of purified Fab' (240 Fab' fragments per 60-fold symmetric virus) for 30 min at 25 °C. Small aliquots of this mixture were applied to holey carbon grids (C-flat). The sample was then flash-frozen by plunging using a Vitrobot (FEI) at 100% humidity and 4°C and using a 2 sec blot time. 1503 images of the specimen were collected at 37,000x magnification and 120 keV on a FEI Titan Krios equipped with a Gatan Ultrascan 4k x 4k CCD camera using the Leginon system (Suloway et al., 2005). The final pixel size was 2.225 Å.

Reconstruction of AAV2-A20 Fab' complex

Appion (Lander et al., 2009) was used for particle picking, CTF estimation, and stack making. Initial particle selection was performed using the difference of Gaussians method (Voss et al., 2009) and particles over carbon were manually deselected. CTF estimation was performed with the ACE (Automated CTF Estimation) software package (Mallick et al., 2005). Images with an ACE CTF estimation confidence value of less than 0.7 were removed from the data set after which a 35,543 particle stack was made with the phases flipped for individual particles according to their ACE-estimated defocus. EMAN (Tang et al., 2007) was used for subsequent refinement

and reconstruction. Refinement first yielded a reconstruction with a resolution of 9.8 Å by the $FSC_{0.5}$ criterion. Inspection of the class averages from the refinement suggested conformational heterogeneity in the Fabs. Subclassification was employed to test for and, as necessary, remove heterogeneity. Multivariate statistical analysis was performed on each class of particle projection following each iteration of refinement. Correspondence analysis was performed with hierarchical ascendant classification on aligned particles in a given orientation. This generated 2-6 subclasses for each projection. Only the subclass correlating best with the projection of the current model was included in the reconstruction for that iteration. After refinement, this procedure resulted in an 8.5 Å resolution ($FSC_{0.5}$) final reconstruction based on 11,898 particles. The whole refinement was repeated using two different starting models: a prior reconstruction of the native virus (O'Donnell, Taylor, and Chapman, 2009); and a model created *de novo* using the EMAN STARTICOS program. These independent refinements converged on the same solution.

Scaling

Correction of the relative magnification proved to be critical. After data collection, it was discovered that the relevant magnification calibration parameter had not been set appropriately in this first structure determined using a newly commissioned microscope. Thus, such a large correction to the relative magnification (1.075) had not been anticipated. Three approaches yielded the same value for the relative magnification. Firstly, a search was made for the best agreement between the

density of the complex and an earlier reconstruction for the native virus (O'Donnell, Taylor, and Chapman, 2009), masking out Fab' density through use of radial cut-offs of 72 & 120 Å. Secondly, a search was made for the highest correlation between the experimental map of the complex, and the local density calculated from the AAV-2 (uncomplexed) crystallographic structure (Chapman, 1995; Xie et al., 2002). Finally, the magnification was least-squares refined by optimizing the agreement of the full atomic model of the complex (see below) and the experimental EM density. Correction of the magnification had a dramatic impact on the flatness of difference maps, and the model-map correlation, improving from 0.2 to 0.7.

Creating a homology model

Initially, the Fab' was modeled using an arbitrarily selected antibody from the protein data bank (PDBid 1A6T), without reference to the actual sequence of A20. This was sufficient to uniquely define the approximate orientation of the Fab' (Figure S2), but a more accurate binding footprint would require a homology model with CDRs of length and conformation appropriate to the sequence of MAb A20. RNA was extracted and cloned from snap-frozen hybridoma cells for PCR amplification using degenerate VH and VL primers and bidirectional cDNA sequencing of the A20 variable domains (Molecular Cloning Laboratories). Modeller (Eswar et al., 2006) was used to create a homology model of the antibody's variable region. 19 structures from the protein databank with a sequence similarity of 58% or greater were used as templates. 1000 models were produced by aligning the templates to the A20 sequence. Of these, the top 100 models had DOPE scores

ranging from -12337 to -12149 and were fitted into the density by Modeller. Of these, the top 10 models had a correlation coefficient of 0.85 - 0.86. The model with the highest combined DOPE score and correlation coefficient was used as the variable domain in an initial homology model for A20. To model the constant region, of known IgG Fab crystal structures, mouse monoclonal antibody 184.1 (PDBid 1osp (Li et al., 1997)) was chosen on the basis of the highest sequence identity.

CDR database models

Alternative models for the variable domain were generated from the Dunbrack database of CDR loop conformations compiled from > 300 non-redundant high resolution crystal structures (North, Lehmann, and Dunbrack, 2011). For CDRs of a given length, conformations fall into a handful of clusters with sequence fingerprints that have been characterized. For A20 CDRs L1 – L3 and H1 - H2, there was a unique sequence match to a database cluster, and for H3, four were possible, two of which could be eliminated later as segments of five or eleven amino acids extended outside the A20 electron density. Median structures coming from the best database cluster for each CDR were spliced together into a single A20 model after superimposing their constant regions. The process was repeated for the alternate H3 conformation. The sequences were then changed to that of A20, and, where necessary, new side chain rotamers were selected to resolve clashes or bring side chains into density. No additional optimization of these database-derived models

was performed. They were used to assess uncertainties in the homology model and in the model-derived footprint.

Docking & Refinement of the homology model

The FAb A20 homology model was fit approximately into a difference map calculated by subtracting a native *cryo*-EM reconstruction from that of the complex. Initial rigid-group conjugate gradient refinement was performed with the Flex-EM option of Modeller-9 (Topf et al., 2008). This revealed two locations where the automatic homology modeling could be improved, residues 209-210 that overlapped with residues 263-264 of AAV-2, and Lys_{69} which extended beyond the difference map and clashed with AAV2 325. Residues 205-216 were remodeled using another high-scoring homolog (model 554) that did not conflict, and an alternate favored rotamer (Dunbrack, 2002) was selected for Lys_{69}.

The structure was further optimized using a new implementation of the real-space refinement RSRef (Chapman, 1995), embedded in CNS (Brünger et al., 1998). From this point, instead of using a difference map, the Fab' model was refined in the presence of AAV-2 into the reconstruction of the complex. Additional refined parameters included the relative EM magnification, and the "soft" resolution limit of a 5^{th} order Butterworth low-pass filter (Frank et al., 1996) applied when calculating the density of the atomic model. The experimental map had been corrected (sharpened) by application of the inverse envelope function with EMB-factor (Fernandez et al., 2008). The low-pass filter allowed the model density to replicate

the resulting smooth, but non-Gaussian attenuation near the resolution limit. In refinement, the squared difference between observed and calculated electron density levels was minimized using all map grid points within 10 Å of any model atom, and considering density contributions from atoms up to 25 Å away. Icosahedral symmetry was imposed as a constraint on both Fab' A20 and the AAV-2 capsid protein. The virus structure was fixed, aligned to the icosahedral symmetry of the map. The Fab' was refined first as a single rigid group, optimizing a weighted (~20:1) sum of the density residual and CNS van der Waals repulsion terms. The Fab' was then split into variable and constant domains for rigid-group positional and group B-factor refinement.

Modeling of the A20 structure led to a 0.8 Å clash at AAV-2 Lys_{258}. This was resolved by choosing a different high frequency rotamer for Lys_{258}. The closest remaining contact was 2.4 Å. No attempt was made at the subtle adjustments needed to resolve contacts that were suboptimal by only ~ 0.5 Å given the resolution of this study. The need for at most subtle adjustment of side chains is consistent with the absence of features in the EM density indicative of significant conformational changes.

Density-based footprint

AAV amino acids contacted by Fab A20 were identified as those with any atom within 4 Å of any homology model atom. An alternative model-independent identification was performed as follows: Between density of the Fab' and AAV-2, a

clear constriction at the periphery of the interface could be used to demark an outline. The contour level of 2.2 σ was used, because it corresponded to the calculated Fab' solvent-excluded volume in a map of the complex segmented in Chimera (Pettersen et al., 2004). On the 2.2 σ contour of the non-segmented map, the boundary between Fab' and AAV-2 was outlined in 3D by connecting saddle points with markers placed interactively using Coot (Emsley et al., 2010). The outline was projected onto the surface of the AAV-2 model using Rivem (Xiao and Rossmann, 2006), identifying AAV-2 amino acids whose surfaces were > 25% covered. A solvent-accessible outline was obtained by extending the markers outwards by the 1.4 Å radius of a solvent probe, and additional potential contact regions were determined using the corresponding 1.1 σ contour level.

RESULTS AND DISCUSSION

Use of Fab' fragments

EM imaging takes advantage of the 60-fold icosahedral symmetry, for which it was important to achieve near saturated binding of equivalent binding sites on the virus. At the required stoichiometries, attempts to image the complex of AAV-2 with intact purified MAb A20 were foiled by aggregation, either when virus and antibody were pre-mixed in solution, or when antibody was added to virus pre-loaded onto EM grids. With papain-digested Fab fragments, prepared the conventional way, neither unstained or negatively stained EM revealed fragments bound to the AAV surface. Even large excesses of Fab failed to compete with intact Mab in dot-blot assays using

biotinylated Mab A20. These data suggested that the papain-digested fragments were not stable in a native fold. By contrast, pepsin-digested Fab' fragments competed with MAb in dot-blot assays, and could be seen decorating the surface of AAV-2 in negatively stained EM images.

Electron microscopy reconstruction

The complex of AAV-2 with Fab' fragments of neutralizing monoclonal antibody A20 (Wistuba et al., 1995) was visualized through *cryo*-EM and reconstructed to a resolution of 8.5 Å ($FSC_{0.5}$) (Figure 17). 11,898 particles from 1066 images were combined in a 3-D reconstruction. Class averages of the complex resolved the variable and constant Fab' domains, with the presumptive variable domain, bound to the capsid surface, better defined than the constant domain (Supplementary Figure 1). The EM reconstruction clearly shows the distinctive features of AAV, previously seen in the uncomplexed virus (O'Donnell, Taylor, and Chapman, 2009; Xie et al., 2002), namely the three-fold peaks, the 5-fold cylinder, 2-fold valley, canyon, and plateau (Figure 18). All of the differences in the capsid density between the complex and native forms could be attributed to differing resolution, i.e. there is no evidence of large scale antibody-induced conformational change in the virus.

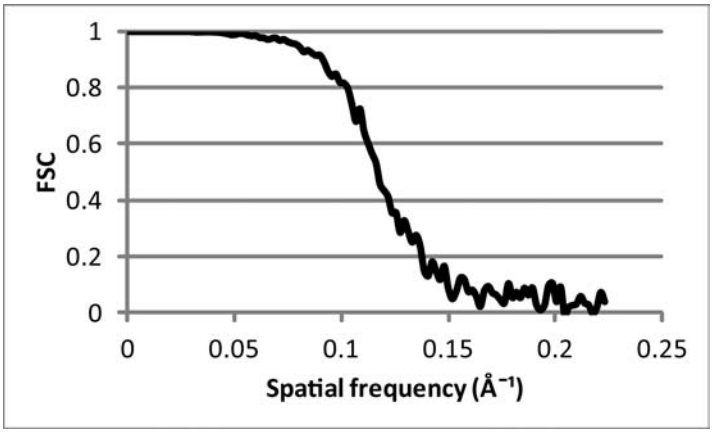


Figure 17: Fourier shell correlation. At a conservative threshold of 0.5, the resolution would be assessed as 8.5 Å. A threshold of 0.143 yields 6.7 Å.

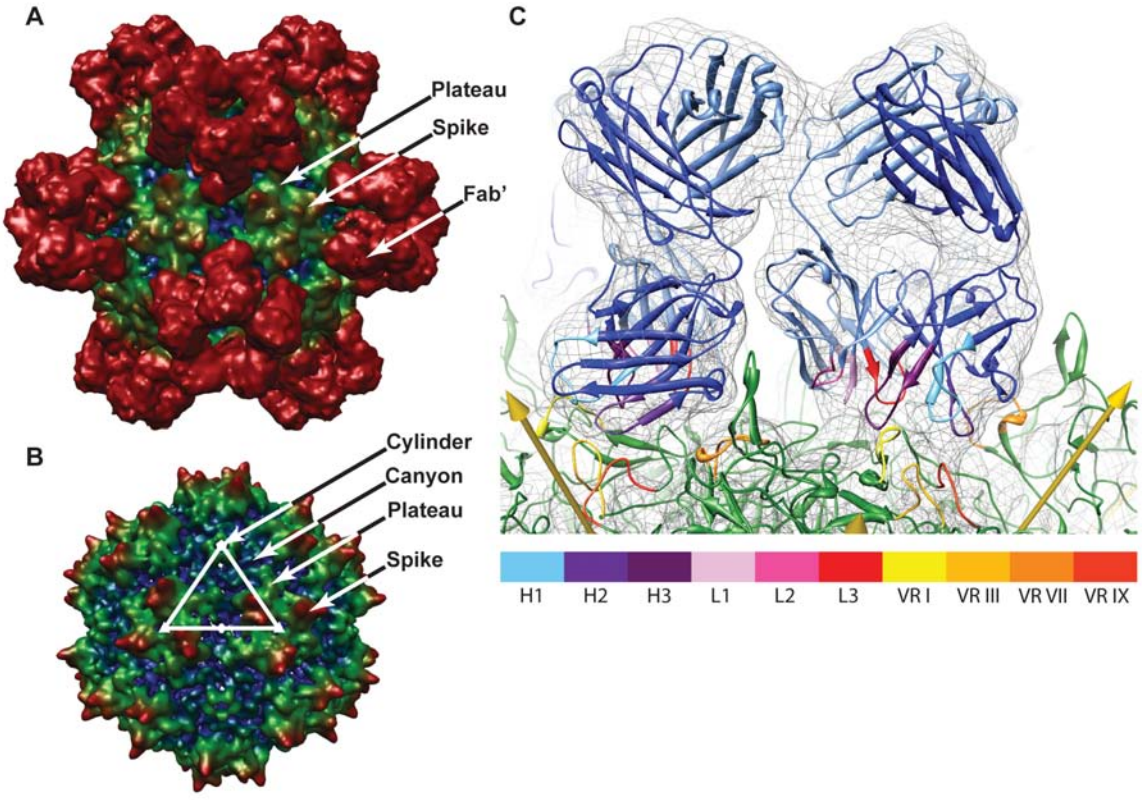


Figure 18. Reconstruction of AAV-2 complexed with A20 Fab' (A) compared to a prior *cryo*-EM reconstruction of native AAV-2 (B) (O'Donnell et al., 2009). The view is along a 2-fold with additional 2-fold axes horizontal and vertical. The asymmetric unit is outlined in panel B with a pentamer at the

5-fold vertex, triangles at the 3-folds and an oval at the 2-fold axis. Reconstructions are colored by radius from the center from blue to red. The canyon surrounding each 5-fold and the depressed area near each 2-fold are blue, the cylinder and plateau are green, and both the 3-fold proximal viral spikes and Fab' are red. C: Fit of the homology model into the density of the *cryo*-EM reconstruction. Two symmetry-equivalent Fab' fragments are shown in blue (pale for the light chain, dark for heavy) with CDRs highlighted according to the color key, H1-3 and L1-3. AAV-2 is shown in green with variable regions VR-I, -III, -VII and -IX, the main CDR contacts, highlighted in color. The view is tangential to the virus surface with yellow arrows indicating symmetry axes, 2-folds on each side, and looking from a 3-fold (front center) towards a 5-fold (out of view). Density is contoured at 2.5 σ. Supplementary Figure 3 shows a stereographic version of panel C, from which the quality of fit can be more readily appreciated.

Homology modeling

A20 is a murine Ig3 antibody whose sequence and 3-D structure have not previously been reported. Using sequences determined here for the variable regions of light and heavy chains (Table 3), Modeller (Eswar et al., 2006) was used to create a homology model of the antibody's variable region. The structure of the complementarity determining regions (CDRs) depends on loop refinement, sometimes a challenging step, so the homology model was checked against the database of CDR conformation based on > 300 antibody structures (North, Lehmann, and Dunbrack, 2011). CDR conformers were used if their database frequency among those with A20-matching sequence motifs exceeded 5% (North, Lehmann, and Dunbrack, 2011). Excluding two conformations of CDR H3 with 5 and 11 C_α outside the A20 density respectively, the frequency-weighted RMS C_α distance between the A20 homology CDRs and matching database conformers was 0.7 Å.

Five of the 6 CDRs were in excellent agreement (RMS C_α = 0.4 Å), with greater uncertainty in H3 (RMS C_α = 1.5 Å). In principle, the homology model, built to fit the *cryo*-EM density (Eswar et al., 2006), should be superior to the un-optimized density-independent cluster structures, and the discrepancy between them provides an upper-limit estimate of the homology model backbone error.

```
Light Chain
                                      CDR1                        CDR2
 1 DIQMTQSSSSFSVSLGDRVTITCKASEDIHNRLAWYKQKPGNAPRLLISGATSLETGVPS 60
      βA    βA'         βB           βC        βC'
                                  CDR3
61 RFSGSGSGKDYTLSITSLQNEDVATYYCQQYWIGPFTFGSGTNLEIK 107
     βD     βE              βF                βG
Heavy Chain
                                CDR1                      CDR2
 1 SDVQLQESGPDLVKPSQSLSLTCTVTGYSITSGYTWHWIRQFPGNKQEWMGYIHFSGYTN 60
        βA              βB              βC          βC'      βC''
                                                CDR3
61 YNPSLKSRVSITRDTSKNQFFLHLNSVTTEDTATYYCARGDYGYEWFTYWGQGTLVTVSA 120
  C''         βD        βE              βF          βG
```

Table 3. Sequence of the A20 variable regions for both light and heavy chains. The complementarity determining regions are annotated for each chain.

Analysis of the 3-D reconstruction

Initial rigid group refinement using RSRef (Chapman, 1995) relieved strain in the inter-molecular van der Waal's energy (by ~20%) with little change in the quality of the fit (CC = 0.775). (Correlation coefficients are calculated for voxels of the map within 10 Å of any Fab' or virus atom.) Least-squares fitting of a Butterworth low-pass filter to the model density lowered the residual difference with the experimental density by 1%, and suggested that the effective resolution of modeling

was slightly lower (10.6 Å) than would be expected of the $FSC_{0.5}$ (8.5 Å), noting that the former reflects the limitations of both experiment and modeling. Splitting the Fab' into variable and constant domains led to only a small improvement in correlation (0.778 to 0.781), but a 40% improvement in intra-protomer van der Waals energy as strain was released. Occupancy of the FAb' refined robustly to 1.0, indicating saturated binding. Refinement of isotropic group atomic displacement parameters (ADP, "B-factors") showed that the variable domain was indistinguishable from the capsid protein (<B> = 23.5 $Å^2$), but the more distal constant domain had an optimal B of ~ 300 $Å^2$. These correspond to rms harmonic displacements of U = 0.3 Å and U = 2 Å for the variable and constant domains respectively. Thus, the variable domain can be considered tightly bound with displacement parameters that reflect the experimental resolution. By contrast, constant domain displacements of 2 Å, and perceptibly more diffuse EM density, are consistent with flexibility at the hinge. A final round of positional refinement converged at CC = 0.787 (Figure 18). From start to end of this refinement the rms coordinate change was 0.8 Å.

Binding site

With virus-Fab complexes, it is usual to analyze the interface indirectly, using a model fit to the Fab density, because the constraints of fitting a known fold, like the imposition of stereochemical restraints in crystallographic refinement, can support analysis beyond the nominal experimental resolution (Rossmann, 2000). It does introduce the errors of Fab homology modeling, which are modest for the backbone

(see above) and depend mostly on uncertainty in choice of side-chain rotamers. It therefore limits analysis to amino acid level, which is nevertheless a potential improvement upon the 8.5 Å primary data. It was cross-checked in two ways (Supplementary Table 1). A model-independent footprint was determined by tracing saddle-point vectors between the virus and Fab densities, contoured to correspond to either the solvent excluded or solvent accessible volumes (Gerstein et al., 2001). Footprints from the solvent-excluded density and homology model are similar (Figure 19), sharing 12 of 16 amino acids, and only 2 homology model contacts fall outside the footprint density contoured to approximate the solvent accessible surface. The impact of homology modeling errors was also assessed by comparison with an alternative model (see above), consisting of spliced loops from Dunbrack's CDR database (North, Lehmann, and Dunbrack, 2011) with matching sequence motifs. 80% of the contact residues were identical. In summary, of 16 footprint residues implicated by the homology model, 12 are robustly implicated by all methods, two are doubtful (Ser_{658} and Thr_{660}) lying outside the density footprint, and a total of 7 additional residues on the periphery are suggested as possibilities by one or more of the alternative approaches. Discussion will be framed around the core consensus footprint, except as explicitly noted.

Residues implicated in the footprint are listed in Table 4 and Supplementary Table 1 and illustrated in Figure 19. The footprint is comprised of multiple peptides contributed by three different symmetry-related subunits, explaining why MAb A20 is specific for assembled capsids (Wobus et al., 2000). It encompasses regions of diverse sequence on the plateau and side of the spike, but also nearly as many in the

more conserved canyon region. The more exposed parts of the footprint contain residues from 4 of 9 "variable regions" (VR) previously noted to have the highest sequence diversity among AAVs (Agbandje-McKenna and Chapman, 2006; Govindasamy et al., 2006). Earlier difficulties in linear epitope mapping are explained by the finding that 6 distinct segments of polypeptide chain contribute to the footprint; mAb A20 binds to a conformational epitope defined by protein tertiary structure, and not a simple linear sequence of amino acids.

The A20 homology model contains a light chain and a heavy chain, with three CDRs per chain. Contacting the AAV plateau are heavy chain residues from CDR1, 2 and 3 (aka H1 – H3) and light chain CDR3 (L3) (Figure 18 & S3). It is the light chain, mostly CDR1 and 3 (L1 & L3) that contact the canyon and two (non-CDR) framework residues from the heavy chain that come close to the side of the spike.

It has been reported that human rhinovirus is neutralized by bivalent antibody attachment in some cases (Hewat and Blaas, 1996; Rossmann et al., 1985). For bivalent attachment adjacent Fab C-termini need to be located within 25-29 Å of each other (Hewat and Blaas, 1996; Hewat, Marlovits, and Blaas, 1998; Thouvenin et al., 1997). For the A20 homology model, symmetry-related C-terminal ends of A20 Fab' are 47 Å apart. The resolution and distinctive features of the reconstruction leave little ambiguity in the orientation of the Fab', with the C-termini of the heavy chain Fab' fragments pointing outwards away from their nearest neighbor and incompatible with bivalent binding of the intact antibody. Thus, although symmetry-related antibody footprints come within 12 Å of each other, only

monovalent attachment would be possible without implausible levels of distortion at the elbow.

MAb A20 neutralizes at a step subsequent to cellular attachment (Wobus et al., 2000). AAV enters cells endosomally, after initial attachment to a heparan sulfate proteoglycan (HSPG) (Summerford and Samulski, 1998), and subsequent binding to one of several possible protein co-receptors (Agbandje-McKenna and Kleinschmidt, 2011). The HSPG binding footprint has recently been mapped by electron microscopy (O'Donnell, Taylor, and Chapman, 2009). Co-receptor footprints are currently unknown. (A site suggested by mutations affecting transduction but not heparan-binding (Lochrie et al., 2006) actually fell within the HSPG footprint (O'Donnell, Taylor, and Chapman, 2009), so the sensitivity of heparan binding assays may not support such specific designations.) Here, we see that the MAb A20 and heparan binding footprints do not overlap. This is consistent with the earlier finding that A20 does not inhibit cellular attachment (Wobus et al., 2000). It remains to be seen whether A20's neutralization results from interference with co-receptor mediated entry, or conceivably some later step. A20 occludes considerably more of the surface than the footprint residues making direct contact and thus could affect other interactions essential to AAV's life cycle. A20's variable domain extends deep into the canyon and is nestled between the 5-fold cylinder and the 3-fold spikes. Although none of the residues of the cylinder and only two of the spike make sub-4Å contacts designated "footprint", a number of residues come within $\sim$ 6 Å, close enough to occlude bio-macromolecules whose interactions may be essential to AAV's life cycle.

	AAV-2				MAb A20
	Residues	**Subunit**	**Variable Region**	**Chain**	**Contact Regions**
Plateau	$_{261}$SSQS$_{264}$	B	I	H	CDR3 + CDR1 (for S260)
	Ser$_{384}$, Gln$_{385}$	B	III	H	CDR3 (for S384) & CDR2 (for Q385)
	Val$_{708}$	A	IX	H	CDR2
	Asn$_{717}$	A		L	CDR3
Canyon wall	Lys$_{258}$	B		L	CDR1/3
Canyon	Asn$_{253}$, Asn$_{254}$	B		L	CDR1 + CDR3 (for N254)
	$_{658}$(S)T(T)$_{660}$	I		L	Framework (for S658 & T660) & CDR1 (for T659 & T660)
Spike	(Glu$_{548}$, Lys$_{556}$)	A	VII	H	Framework

Table 4. Regions of AAV-2 contacted by MAb A20. Residues were identified that fell within 4 Å of the fitted Fab' homology model. Most of the footprint is validated by comparison to two independent assessments (Table S1). Parentheses denote residues on the periphery that were not implicated by all three approaches.

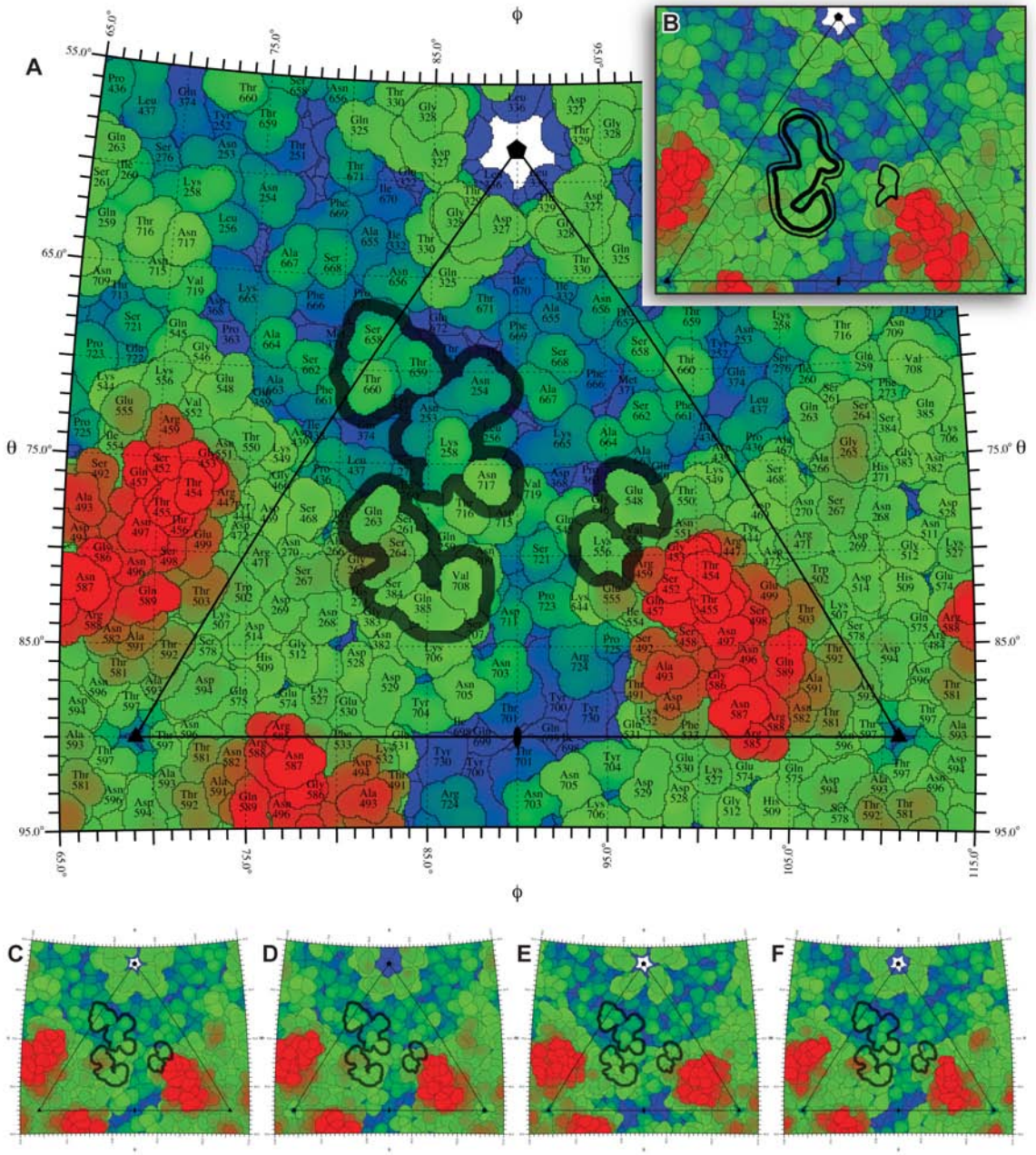


Figure 19. Stereographic polar projections of the Fab' A20 footprint (outlined in black) onto the AAV asymmetric unit, outlined by the triangle (see Figure 1B). Color indicates distance from the viral center, from blue (105 Å) to red (140 Å). (A) Footprint determined from a homology model fit to the EM density projected onto the surface of AAV-2; (B) Footprint determined directly from the 8.5 Å density, contoured at a level to correspond to the solvent-excluded volume (thick line) or solvent-

accessible surface (thin line); (C – F) Projections of the AAV-3B (C), AAV-4 (D), AAV-6 (E), and AAV-8 (F) structures onto which is overlaid the AAV-2 Fab' footprint from panel A.

Comparison with autonomous parvoviruses

There are not yet other *Dependovirus*-antibody complexes with which to compare, but there are several *cryo*-EM structures from the autonomous parvoviruses of the sister *genus Parvovirus*. The greatest similarities are with antibodies that bind to epitope B of the closely related pair, canine parvovirus (CPV) and feline panleukopenia virus (FPV) (Hafenstein et al., 2009; Wikoff et al., 1994). The AAV-2 A20 binding site therefore contrasts with the binding site of MAb B7 of minute virus of mouse (MVM) which is directly above the 3-fold axis of symmetry (Kaufmann et al., 2007). The surface topology of the autonomous viruses at the B7 site differs from AAV with loops from adjacent subunits coming together to form more of a single elevated "massif". AAV has 3 distinct protrusions, between which are valleys that meet at the 3-fold, so the B7 site is less accessible. However, some common themes are emerging. None of the complexes, including AAV-2/A20, are compatible with bivalent attachment, and several of the binding sites: MVM B7, CPV/FPV B-site and AAV2 A20, are at interfaces where loops from neighboring subunits come together to define the viral structure.

Cryo-EM-defined footprints for six CPV/FPV MAbs overlap at site "B", comprised mostly of loop 3, a region implicated by mutagenesis to be antigenic (Hafenstein et al., 2009; Wikoff et al., 1994). CPV residues 298 to 302 superimpose over the AAV-2 A20 footprint if aligned by icosahedral symmetry. The B site and VR1 of the AAV-2

A20 footprint are located on the plateaus of their respective viral surfaces. In CPV/FPV, bound antibodies overlap with the binding site of the transferrin receptor, blocking both viral attachment and subsequent entry steps (Hafenstein et al., 2007), but the steps of AAV entry and mechanisms of A20 neutralization are not as well characterized. The considerable similarities appear to result from convergent evolution, because AAV's VR1 does not correspond in primary structure with loop 3 of CPV/FPV. VR1 is in the loop between β-strands B and C, which in CPV/FPV is pushed up from the plateau as part of the 3-fold massif, and comprises epitope A, to which a distinct 3 MAbs bind (Hafenstein et al., 2009). Paradoxically, epitope B (on the plateau of CPV and FPV) is formed by residues from the loop between strands G and H whose equivalents are at the exposed tip of the 3-fold protrusions in AAV. In the autonomous viruses, the plateau extends further out from the 3-fold axis, and the footprints of B-site antibodies all fall within the bounds of the plateau. With AAV-2, the topography is more varied, and for the first time we see an Fab binding site extend over the edge of the plateau into a more prominent and less exposed canyon.

Epitope

As the first AAV neutralizing monoclonal antibody (Wobus et al., 2000), MAb A20 has been the subject of repeated epitope mappings using multiple approaches (*vide infra*). The AAV-2 crystal structure (Xie et al., 2002) showed that not all regions implicated by peptide scanning (Moskalenko et al., 2000; Wobus et al., 2000) could fall within a typical antibody footprint, or, indeed were near the outer surface of the

virus. Thus, structure has already helped integrate the molecular data, complementing scanning approaches and their known limitations (Van Regenmortel, 1992). Visualization of the A20-AAV2 binding interface provides an additional perspective. Exact consistency with molecular approaches should not be expected, due to ambiguities in each approach, including the potential for wider conformational change in escape mutants, and limited precision in structural studies at 8.5 Å resolution (see above). Nevertheless, integration of a physical epitope from EM visualization with the functional molecular approaches can provide a more robust understanding of immunogenicity.

There has been wide variation in surface sites implicated in A20 binding, but the plateau has been implicated more often than other regions. Single residue substitutions at 64 surface-exposed sites revealed a cluster that affected A20 binding or neutralization: Q263, S264, S384, Q385 and V708 (Lochrie et al., 2006). These are all within the EM footprint, and are joined in close proximity by the sites of peptide insertions after S261, A266, N244 and S247, the first three of which inhibited A20 binding (Girod et al., 1999; Wu et al., 2000). Thus, from this molecular data and our structure, a consistent picture emerges that an important part of the A20 epitope is the plateau where the VR-I and VR-III loops from one subunit and VR-IX from a neighbor pack together.

The EM reconstruction shows the A20-AAV2 binding surface extending into the canyon, a region that was not tested by Lochrie *et al*. (2006) or other studies. To the EM footprint are added S261, S262 and N717 on the edge of the plateau, K258 on

the wall, then N253, N254, T659 and possibly S658 and T660 in the canyon. As in the complex of human rhinovirus (HRV) 14 with Fab17-1A (Smith et al., 1996), the CDRs of MAb A20 penetrate a less exposed, more conserved region of the surface, although in AAV, the region has not yet been associated with any viral function. Extending the binding site into the canyon approximately doubles the size of the footprint implicated by Lochrie's cluster of plateau mutants.

The core residues of the footprint are implicated both in model-based calculation of contacts, and in projections of experimental Fab' density onto AAV-2 that are not dependent on modeling the Fab'. A few residues on the periphery are less certain, because their designation is sensitive to the choice of solvent probe or density contour level. E548 and K556 come within 4 Å of modeled A20 atoms, but are connected by density only at lower contours corresponding to the solvent accessible surface. Mutation of E548 has been seen to yield a small 3-fold resistance to A20 neutralization (Lochrie et al., 2006) or, like K556 (*ibid*), to have no impact upon binding (Wu et al., 2000). It seems likely that the observed proximity of the Fab' to two residues on the side of an adjacent AAV-2 spike is incidental and not critical to the binding interactions. Indeed, the spike residues are apparently not part of the epitope, because the contact would be with conserved framework residues towards the side of the immunoglobulin domain, not CDR loop residues usually associated with antigen recognition.

Distinctiveness of AAV-2 and AAV-3B

AAV-2 and AAV-3B are the only serotypes which are bound strongly by A20. With only limited success, a search was made for simple distinguishing characteristics of the footprint that would be conserved between AAV-2 & AAV-3B, but distinctly different for the other serotypes that are not bound. At 8.5 Å resolution, side chain conformation is not defined for the Fab', so complementarity in the atomic properties of the interacting surfaces cannot be investigated. However, the Fab' footprint makes it possible to compare the relevant regions of the surfaces in the 2.5 to 3.0 Å crystal structures of the different serotypes.

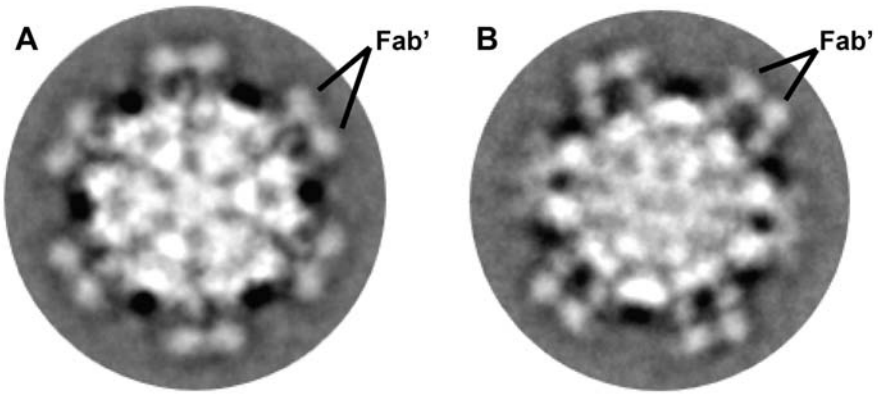
In sequence, AAV-3B differs from AAV-2 only at three binding site residues, all on the edge of the A20 footprint: S658 is replaced by P659 in the canyon, while E548 and K556 of AAV-2 are replaced by T549 and N557 in AAV-3B on the side of the spike (Figure 19). Lochrie *et al*. (2006) found that A20-binding is not lost upon conservative polar mutations of E548 and K556, and that E548A mutation resulted in only a 3-fold decrease in A20 neutralization. The contact at E548/K556 in the AAV-2 complex is glancing. Density bridges only when contoured to correspond with the solvent accessible, but not solvent excluded surface (Supplementary Table 1), and the contact is with a conserved part of A20, not a CDR, and may therefore be incidental. Thus, the binding of AAV-3B by A20 can be understood in terms of conserved sequence within the core of the footprint, with differences restricted to the periphery at sites that have at most modest impact on binding (Lochrie et al., 2006).

The greater challenge is in understanding what distinguishes AAVs that are not recognized by A20, some of which are also highly conserved. Most of the differences in surface topology are outside the binding site. Within the canyon, the surfaces of AAV-2, -3B, -4, -6, and -8 are all very similar, consistent with their high sequence conservation, and differences on the plateau are subtle and not obviously correlated to A20 binding. For example, AAV-6 (non-binding) differs from AAV-3B less than does AAV-2 (both bound by A20). Near the A20 site, the sequences of AAV-6, -1 and an AAV-2/-1 chimeric construct (Hauck and Xiao, 2003) are similar, differing from AAV-2 with a threonine insertion after S264 and/or a Q263A substitution. There is not a substantial change in surface topology that would create a clear steric conflict with A20 (Figure 19 A & C), but rather the surface is smoothed out and subtly changed in polarity. The A20 binding site on AAV-2 is comprised of predominantly polar residues, with the exception of V708, but there is not a pattern of amino acid polarity or electrostatic charge in the local surfaces that clearly segregates AAV serotypes that are bound by A20 from those that are not. At this point, one can conclude only that A20 binding can be determined by differences as subtle as between AAV-2 and -6. Presumably, it depends upon optimal complementarity of the viral and antibody surfaces leading to a suitably favorable binding free energy, but understanding such intricacies require precision in loop and side chain structure that are not available for the 8.5 Å Fab' homology-based model.

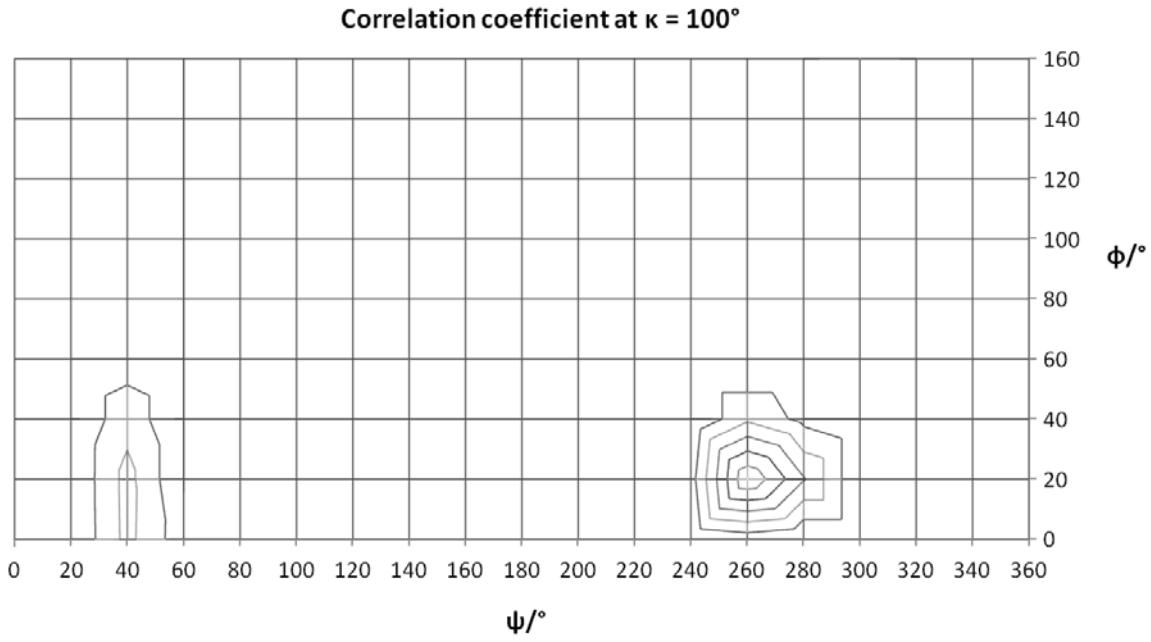
For future efforts to engineer gene therapy vectors with reduced susceptibility to immune neutralization, the reported structure is encouraging. The footprint is larger than previously anticipated, providing a wider choice of residues to mutate

for immune neutralization escape. Relatively modest changes can affect antibody binding, and there is no overlap for binding sites of receptor HSPG and this model antibody, increasing the prospects for engineering the vector without disrupting cell attachment.

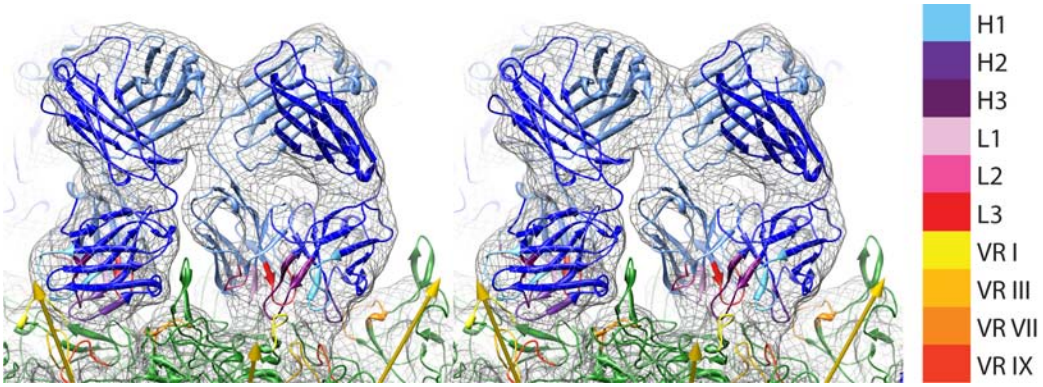
SUPPLEMENTARY DATA



Supplementary Figure 1. *Cryo*-EM imaging of AAV-2 complexed with A20 Fab'. Class averages projected along 3-fold (A) and 2-fold (B) axes respectively, in which the variable and constant domains can be resolved most easily where they emanate radially in a near-equatorial plane. In panel A, viewed down a 3-fold, 6 clusters of 5 Fab' are near-equatorial . In panel B, viewed along a 2-fold, 4 clusters are in-plane and 2 are less well resolved, because we are looking at the superimposition of clusters pointing 30° above and below the plane.



Supplementary Figure 2. Search for the orientation of the Fab' in the *cryo*-EM reconstruction. The structure of a representative Fab (PDB ID 1A6T) was rotated throughout spherical-polar space in 20° increments to get a preliminary model orientation prior to homology modeling. Correlation coefficients above 0.5 are shown (increments of 0.05) for the match of model and experimental density. On this section, the highest and third highest peaks are visible at 0.75 and 0.57. Peaks 2 – 4 at correlation coefficients of 0.62, 0.57 and 0.47 are pseudo-symmetrical solutions related to the correct solution by 2-fold rotations about the principal axes of the approximately ellipsoid Fab' density. Peaks 3 & 4 would implausibly point the CDRs away from the virus. Clear distinction between peak 1 and these incorrect solutions provided confidence that the correct orientation could be found even with a model that only crudely represented Fab' A20.



Supplementary Figure 3. Stereo pair, showing the fit of the homology model into the density of the *cryo*-EM reconstruction. Two symmetry-equivalent Fab fragments are shown in blue (pale for the light chain, dark for heavy) with CDRs highlighted according to the color key, H1-3 and L1-3. AAV-2 is shown in green with variable regions VR-I, -III, -VII and -IX, the main CDR contacts, highlighted in color. The view is tangential to the virus surface with yellow arrows indicating symmetry axes, 2-folds on each side, and looking from a 3-fold (front center) towards a 5-fold (out of view). Density is contoured at 2.5 σ.

			Density	Homology model	CDR database
			Solved excluded (●) & accessible (○) volumes		*Most likely cluster (●) & alternate H3 (○) loops*
Plateau	Ser	261	●	●	●
	Ser	262	●	●	●
	Gln	263	●	●	●
	Ser	264	●	●	●
	Gly	265	○		●
	Ala	266			○
	Ser	384	●	●	●
	Gln	385	●	●	●
	Ser	707	○		
	Val	708	●	●	●
	Asn	709	●		
	Thr	716	●		●
	Asn	717	●	●	●
Canyon wall	Lys	258	●	●	●
	Gln	259	●		
	Ile	260	●		
	Tyr	272			○
Canyon	Thr	251	○		●
	Tyr	252	●		
	Asn	253	●	●	●
	Asn	254	●	●	●
	Leu	256	○		
	Ser	276	○		
	Ser	658		●	
	Thr	659	●	●	●
	Thr	660		●	●
Spike	Glu	548	○	●	●
	Val	552	○		
	Lys	556	○	●	●

Supplementary Table 1. AAV-2 residues potentially within the MAb A20 footprint were identified three ways. (1) Those > 25% covered by Fab' *cryo*-EM density: density was contoured to give a volume equal to the solvent excluded (approx. van der Waals) volume or to the larger solvent accessible volume outlined by the center of a water probe rolled over the solvent excluded surface (Gerstein et al., 2001); (2) Those with atoms within 4 Å of atoms in an A20 homology model fit to the *cryo*-EM density; (3) Those with atoms within 4 Å of a model spliced together from the CDR loops of appropriate sequence motif from Dunbrack's data base of CDR conformation (North et al., 2011). Open circles denote additional residues implicated by an alternative CDR H3 conformer that cannot be excluded by sequence, but is 6-fold less frequent in the database, and, with six Cα clashing with AAV-2, is unlikely to be the relevant conformation. Usually, a fit homology model is considered to provide the most accurate footprint, but is subject to

modeling errors, particularly in side chain rotamers. The membership of a core 12 residues in the footprint is validated by consistency between the three approaches. Residues implicated by a subset of approaches lie at the periphery of the footprint.

AFTERWORD: AAV-2 IN COMPLEX WITH ANITBODY A20

INTRODUCTION

This chapter completes several lines of research that followed naturally from the structure of the AAV-2/Fab' complex presented in the previous chapter, but would not be appropriate for publication in a peer-reviewed journal. First, a return is made to the competition ELISA experiments of "Assessment of A20 Fab Functionality: Competition Dot-Blot Assay" to show that, unlike the Fab fragments, Fab' fragments compete for binding with whole MAbs. There follow attempts to analyze the viral binding site of FAb A20, searching for physio-chemical properties that are shared between cross-reactive serotypes, but distinct for other serotypes, attempting to understand the basis of binding specificity. Such attempts depend on the homology model for Fab A20, which, it was shown, was reliable at backbone level, but of unknown accuracy in side chain configurations that underlie atomic interactions. Thus, this represented an exhaustive / extrapolative analysis of the available structural data.

METHODS

Assessment of Fab' Functionality

Dot blots were used to assess A20 Fab' biological functionality. 0.57 µg of AAV-2 was pre-incubated with varying amounts of Fab' overnight at 4° C and immobilized on a

nitrocellulose membrane. Immobilized virus on the membrane was blocked with 5% BSA in TBS-T. 2.5 ng of Purified A20 MAb in 0.1% BSA/TBS-T was incubated with the immobilized virus for 30 min at room temperature. The sample was then blocked in 5% BSA in PBS-T for 1 hr and subsequently incubated with 0.05 ng/µL of biotinylated A20 in TBS-T for 1hr. The sample was washed and incubated with a 1:200 dilution of alkaline phosphatase conjugated streptavidin (Pierce). After washing, the sample was then detected with 1-Step NBT/BCIP (Thermo Scientific) detection reagent until desired intensity was reached.

Electrostatics of the A20 Fab' Binding Site

APBS was used to calculate the electrostatic potential of the A20 homology model, AAV-2, AAV-3b, AAV-4, AAV-6, and AAV-8 (Baker et al., 2001). APBS uses the Poisson-Boltzmann equation to model the electrostatic properties with a second-order nonlinear elliptic partial differential equation. APBS determines the electrostatic potential based on the dielectric properties of the system, the solution's ionic strength, and the atomic partial charges of the solute. The electrostatic potential was calculated with 150 mM solvent at 25 °C. A 1.03 Å coarse grid was used to approximate the electrostatic potential of the entire virus. A fine grid size of 0.37 Å was used within the asymmetric unit and assessed areas (Figure 21). Imaging was performed with Chimera (Pettersen et al., 2004).

Atomic Solvation Parameters of the A20 Fab' Binding Site

Values for the solvation free energies of each residue of the A20 homology model, AAV-2, AAV-3b, AAV-4, AAV-6, and AAV-8 were taken from Eisenberg and McLachlan and normalized on a 0 to 1 scale (Eisenberg and McLachlan, 1986). Stereographic projections were created with Rivem (Xiao and Rossmann, 2006) where residues were shaded according to their normalized atomic solvation parameter.

RESULTS

Assessment of Fab' Functionality

The biological functionality of A20 Fab' was assessed via a competition dot-blot assay in an attempt to use A20 Fab' to compete off A20 monoclonal antibody. Various amounts of A20 Fab' were pre-incubated with AAV-2 before immobilization in an attempt to block exposed A20 epitopes. In the case that A20 epitopes were blocked by bound A20 Fab' (indicating formation of Fab-AAV-2 complex), these complexes would exhibit reduced signal when incubated with biotinylated A20 primary antibody.

Experimental results indicated there was a significant reduction in intensity at higher Fab' concentrations, suggesting that Fab' competes off A20 MAb (Figure 20). Note that when this assay was repeated with Fab, a reduction in intensity was not seen corresponding to higher Fab' concentrations (*vide supra*).

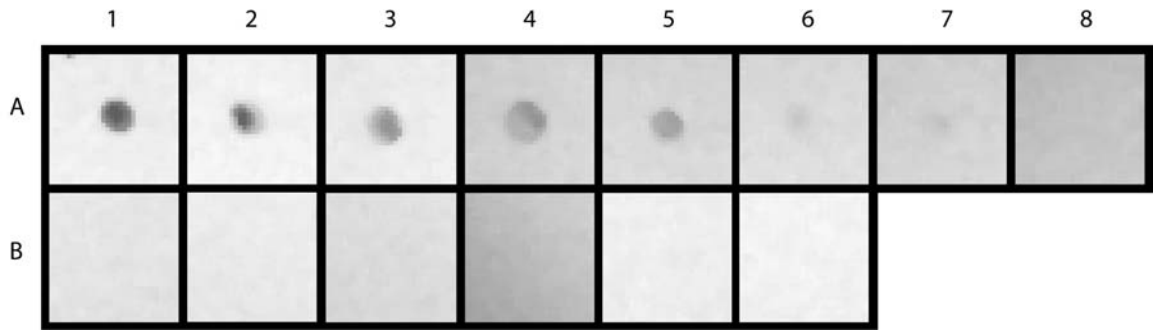


Figure 20. Dot-blot of AAV-2 and varying quantities of A20 Fab' immobilized on a nitrocellulose membrane and incubated with biotinylated A20 primary antibody. AAV-2 was immobilized on nitrocellulose and pre-incubated without A20 Fab' and with AAV-2 (A1), with a 300:1 stoichiometric ratio of A20 Fab' to A20 MAb (A2), a 500:1 ratio (A3), a 700:1 ratio (A4), a 1000:1 ratio (A5), a 5000:1 ratio (A6), a 10000:1 ratio (A7), and a 20000:1 ratio (A8). As a control, unbiotinylated A20 was used as a detecting antibody and incubated without A20 Fab' and with immobilized AAV-2 (B1), with a 300:1 stoichiometric ratio of A20 Fab' to A20 MAb (B2), a 500:1 ratio (B3), a 700:1 ratio (B4), a 1000:1 ratio (B5), and a 5000:1 ratio (B6).

Electrostatics of the A20 Fab' Binding Site

While there is modest inter-serotype variation in the electrostatic potential in the AAV-2/A20 footprint, neither viral nor antibody interface surfaces are highly charged, and there is not strong complementarity between them (Figure 21). In this respect, the AAV-2/A20 interaction is quite different from that of Fab17-IA with human rhinovirus 14, in which strong and specific electrostatic interactions are found (Liu et al., 1994).

With respect to complementarity, the region of the footprint extending into the canyon has an electropositive region which slightly overlaps with an electronegative

region of the homology model. Residue E548 at VR VII on the spike is proximal to a serine on the homology model. The footprint at VR I and III consists exclusively of polar, uncharged serines and glutamines. At VR I, these are proximal to glycines and tyrosines on the homology model. Homology model residues proximal to VR III are tyrosine and phenylalanine. VR IX consists of a hydrophobic valine, which is proximal to a serine on the homology model.

With respect to electrostatic variations between serotypes, AAV-3b unsurprisingly maintains similar electrostatics to AAV-2 due to its high sequence similarity. The other serotypes are slightly more dissimilar, but not in a singularly identifiable way.

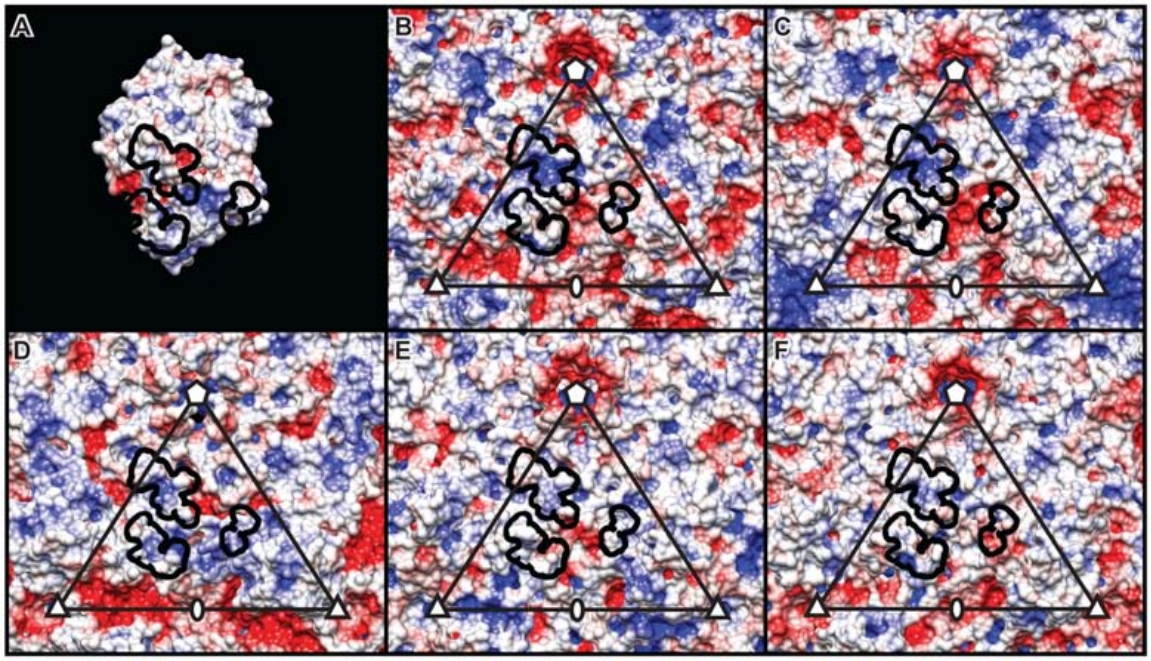


Figure 21. Electrostatic potential projected onto the molecular surfaces of A20 (A) and known AAV structures: AAV-2 (B), AAV-3b (C), AAV-4 (D), AAV-6 (E), AAV-8 (F). The electrostatic potential was calculated with APBS (Baker et al., 2001) and projected onto the surface with Chimera (Pettersen et al., 2004) on a color scale of –5 k_BT/e_c (red) to +5 k_BT/e_c (blue). The triangle shows one of the 60

icoshedrally symmetric areas of the virus, viewed along a 2-fold towards the center as in Figure 2, with the AAV-2/A20 footprint outlined in black. The outline is overlaid on the surface of the Fab' which is viewed from the opposite direction to highlight the corresponding binding surface.

Atomic Solvation Parameters of the A20 Fab' Binding Site

The A20 footprint on AAV-2 is comprised of polar residues in the canyon (N253, N254, N258, N717, S658, T659 & T660). The footprint on the plateau presents a similarly polar surface comprised of multiple serine and glutamine side chains, with the exception of V708, which renders hydrophobic the part of the footprint closest to the 2-fold axis. The footprint on the side of the spike at E548 and K556 presents a mostly polar surface. As for residues of the A20 homology model that interact with these viral surfaces, the residues are of mixed type, with a modest preference for polar.

On projection of atomic solvation parameters (data not shown) or side chain solvation energies (Figure 22) (Eisenberg and McLachlan, 1986; Fauchere and Pliska, 1983) to the viral surface, one is first struck by the level of similarity in polarity of the corresponding surfaces in homologous AAV serotypes, through much of the canyon and plateau regions of the footprint. The side of the spike has the highest variability both in sequence hydrophobicity and in local surface topology for residues that are proximal to A20 if not in direct contact. For example, AAV-8 has an alanine in the position of contact residue E548, but AAV-4 and AAV-6 have a glycine. However, from the E548A mutation in AAV-2 that decreases binding 3-fold (Lochrie et al., 2006), we know that this site alone is not responsible for ablation of binding in

AAV-4, -6 & -8. On the plateau, the lone hydrophobic residue, V708 is bound in a region of A20 with mixed polarity (H54, F55, S56, and Y58). At this resolution, we do not see V708 bound within an A20 hydrophobic pocket that could endow steric specificity, but the presence of a hydrophobic residue at this position distinguishes serotypes recognized by A20 (Val in AAV-2 & -3B) from most that are not (Thr, Asn, Gln or Thr in AAV4, 5, 7-12, but Ala in variants AAV-1/6). 10-fold reduction in binding and 220-fold reduction in neutralization upon V708K mutation in AAV-2 (Lochrie et al., 2006) is consistent with hydrophobicity at this site being one of several partial determinants of strong A20-binding.

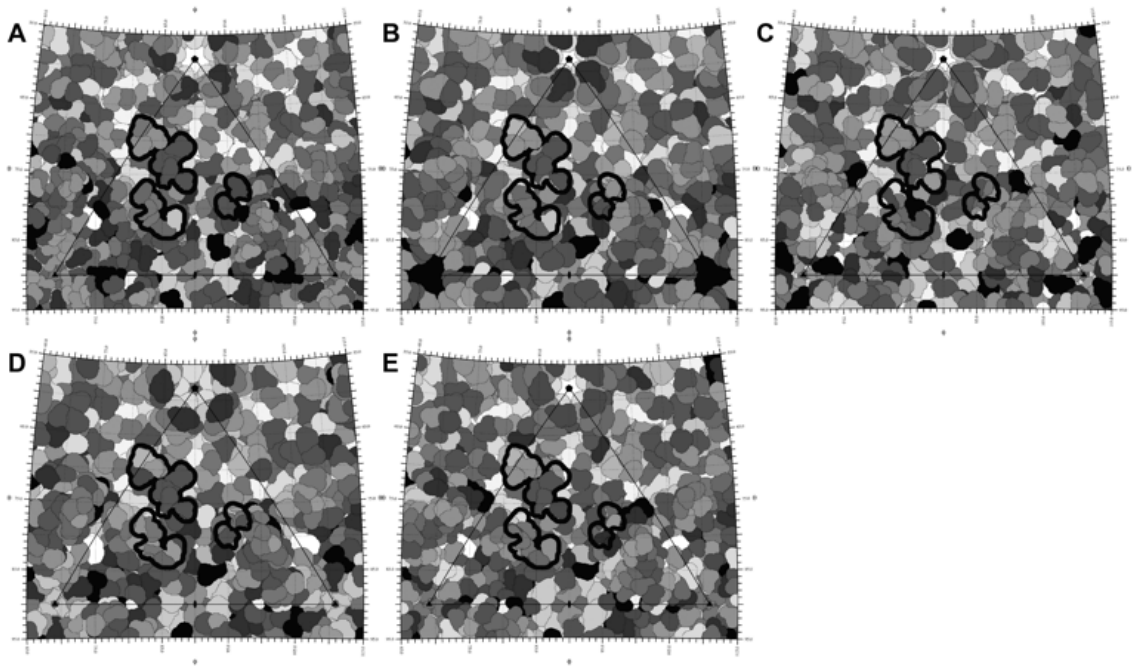


Figure 22. Shading according to amino acid energies of transfer is used to distinguish polar (dark) from hydrophobic (light) regions on the surfaces of AAV-2 (A), AAV-3b, (B), AAV-4 (C), AAV-6 (D), and AAV-8 (E) using the same stereographic polar projection of the asymmetric unit of AAV as shown

in supplementary figure 2. The antibody binding footprint for A20 against AAV-2 is outlined in black on all representations.

DISCUSSION

A relatively high ratio of A20 Fab' to A20 MAb was required to significantly inhibit binding of A20 MAb. However, it is not uncommon for Fabs to have significantly lower binding affinities than their intact IgG counterparts (Alam et al., 2007). The removal of the adjacent Fab and Fc domains of the antibody may have a destabilization effect on the remaining Fab (or Fab'). Papain also digests a small portion of the Fab peptide beyond the nearest cysteine in the hinge region, whereas pepsin digests the Fab' on up to the nearest cysteine in the hinge. It is possible that the digestion of additional peptides severely inhibits the ability of the Fab to retain its binding specificity.

There is not obvious one-to-one complementarity between A20 and AAV-2 at residue level with regards to electrostatic specificity or side chain solvation energies. It is possible that the lack of complementarity is due to unreliability of the side chain rotamers of the homology model, which would only be apparent at residue-level resolution. However, there is also no a single distinguishing feature of AAV-2 and AAV-3b that segregates A20-recognized serotypes from those that are not. Comparison of serotype differences does not rely on the relatively lower-resolution homology model, and instead is dependent on the high-resolution crystal structures of the virus serotypes.

In retrospect, this is perhaps to be expected. Mab A20 is a mouse monoclonal to which AAV, in nature, has not been exposed. Thus, A20 has not itself been driving the evolution of AAV serotypes. The level of homology between A20 and human antibodies would not be high enough to see the evolutionary effects of escape mutations to natural human antibodies reflected in the detailed interactions between human AAV and a mouse monoclonal. What we can infer from the current structure is that A20 recognition is determined by the cumulative effect of several relatively subtle modulations in otherwise quite conserved surface properties.

REVIEW AND FUTURE DIRECTION

REVIEW

To the best of our knowledge, we have determined the first structure of an AAV-antibody complex. The complex structure was acquired with A20 monoclonal antibody, which is the best characterized antibody against AAV serotypes (Girod et al., 1999; Lochrie et al., 2006; Moskalenko et al., 2000; Shi, Arnold, and Bartlett, 2001; Wobus et al., 2000; Wu et al., 2000). Additionally, the A20 antibody is thought to be a potential model for human neutralizing interactions (Lochrie et al., 2006).

This research was more challenging that was originally anticipated due to several unforeseen factors. Preparation of A20 was hampered by precipitation at low temperatures and susceptibility to reduction by cysteine. Sample preparation was further hampered when four different assays of A20 Fab functionality suggested that the purified A20 Fab did not strongly bind to AAV-2, which ultimately necessitated the preparation of A20 Fab' through a different methodology. After the determination of the A20-AAV-2 structure, it became apparent that a number of tools were not optimized for the purposes of this research including scaling electron density maps, model fitting, assessing the accuracy of the orientation, and robustly assessing the interactions of the complex.

The reconstruction of A20 antibody bound to AAV-2 shows that the binding footprint of A20 extends from the plateau to the spike in one direction and into the

canyon in the other. The footprint does not overlap with the previously implicated binding site of HSPG, the site of primary attachment to cells (O'Donnell, Taylor, and Chapman, 2009). The plateau and the spike had been previously implicated by single-site mutations, however Lochrie et al. performed no single-site mutations within the canyon (Lochrie et al., 2006). This research adds several amino acids to the A20 footprint including several in the canyon. Previously implicated E548 (Lochrie et al., 2006) was found to not bind to the A20 Fab' hypervariable loops after investigation via homology modeling, instead binding to the more conserved antibody domain framework, and is thus not considered to be within the A20 footprint proper. Thus, the A20 Fab footprint is clarified from previous studies and the Lochrie et al. footprint is approximately doubled. The larger footprint is promising for the engineering of gene therapy vectors which escape antibody binding, due to the larger range of possible mutations than were previously implicated.

Additional assessment of the electrostatic surface and the atomic solvation parameters of the binding site show that there is no single distinguishing characteristic between AAV serotypes that bind to A20 and those which do not. Therefore, the current structure suggests that A20 recognition is moderated by several relatively small modulations in surface properties at the binding site.

METHODS DEVELOPMENT

MODELLER was used to create a homology model of A20 Fab' which was then refined by rigid fitting of the two independent Fab' domains into the electron density map with RSRef. The resultant model gave an excellent fit to the electron density and the backbones of the hypervariable loops correlated well to the expected cluster medians. However, rotamers which clashed with intermolecular van der Waals (vdW) radii needed to be manually refined and the side chains of the homology model were not expected to be as reliable due to the lack of a high-resolution density map.

Side chains are inherently difficult to model because they require a high-resolution density map and are often times inherently flexible. In the case of complexes, the modeling of side chains is a necessity if intermolecular interactions are to be assessed, even though only a limited resolution map is available. Modeling side chain rotamers with moderate-resolution maps relies on relatively subjective input from the investigator, where the researcher must manually relieve vdW clashes and choose new rotamers for those that extend outside of the electron density map. The methodology of side chain modeling could be improved by limiting the subjectivity of the process. In the following paragraphs, some ideas are proposed to take advantage of information on the likelihood of different side chain orientations that can be unlocked from the PDB, which could be used as an additional restraint for models with unknown side chain conformations.

Side-chain rotamer frequencies have been extensively investigated (Dunbrack Jr. and Cohen, 1997; Gerstein and Levitt, 1998; Lovell et al., 2000). It would be possible to improve the A20-AAV-2 model by retaining a probability-weighted ensemble of side chains using rotamer distributions from the PDB, avoiding the subjective selection of a single rotamer when data are not of high enough resolution to distinguish the different possibilities. Assessing the side chain rotamer probability of a model would allow an additional restraint to be applied to the complex structure and relieve investigator subjectivity. This would be beneficial in many cases where a model is available, but a high resolution density map is not available for precise side chain fitting.

During density-guided docking of the antibody to the virus, the distributions of rotamers could be iteratively improved. This would only need to be done for surface residues at the intermolecular interface. (Side chains buried within each rigid unit are not changing their interactions on rigid-body fitting.) For residues with rotamers that clash, each cycle of refinement would remove the worst overlapping rotamers for each side-chain. A probability-weighted refinement would be used for those that remain, renormalizing the probabilities to total 1.0.

A model produced in this manner will possess side-chain rotamer orientations that are largely independent of the original model. The side-chain rotamers assigned in this way are only biased by the initial backbone structure, the side-chain rotamer libraries, and vdW clashes of nearby side-chains.

The likelihood-estimated footprint may also be used to identify possible alternative surface topologies of the original homology model. Additionally, analyses of electrostatic potential and atomic solvation parameters can be applied to each individual model which represent the likelihood-estimated footprint and then averaged to acquire an electrostatic potential or atomic solvation surface which is representative of the likelihood-estimated footprint.

BIOLOGY

When a patient has neutralizing antibodies against a gene therapy vector, the efficiency of a gene therapy vector is significantly decreased (details below). Neutralizing antibodies to AAV-2 have been found in up to 60% of the U.S. population (Calcedo et al., 2009). In cases where antibodies to AAV-2 are not present, re-administration of therapies has been necessary for cystic fibrosis and hemophilia treatments (Flotte and Carter, 1998; Kay et al., 2000). The vector's efficiency is reduced when re-administered due to patients develop neutralizing antibodies from the first administration (Halbert et al., 2000; Moskalenko et al., 2000; Peden et al., 2004). It is therefore necessary to create a pool of AAV serotypes that are antigenically distinct so that initial and subsequent gene therapy treatments may be administered without an unnecessary reduction in efficiency from the antigenic response of the patient.

Escape mutants are antigenically distinct variations of a virus that escape neutralization by the immune system. There are two approaches to creating escape

mutants that may be especially viable. An adaptive natural selection approach may be used, with selective pressure from monoclonal or polyclonal antibodies, and accelerated with the infusion of increased mutagenesis from an XL1-Red strain of E. coli. This approach currently needs improvement and an alternative method of creating a comprehensive pool of AAV-2 single-site mutants, such as error-prone polymerase chain reaction (PCR), should be considered (*vide infra*).

It is also possible to take a rational design approach that relies on the structural characterization. In this thesis research we have begun to characterize the antigenic surface of AAV-2 from a structural perspective. The binding site of A20 that has been determined helps identify one major immunodominant region on the AAV-2 capsid surface. For canine parvovirus (CPV), the major antigenic epitopes have been shown to be approximated by the subset of amino acids collectively bound by each of 3-5 antibodies that bind to a common epitope (Hafenstein et al., 2009). Extrapolating from studies of CPV, it is expected that two to three major immunodominant regions are present on the AAV-2 surface. Therefore, it would require the structures of 5-15 additional antibody-virus complexes to determine which amino acids are immunodominant. By elucidating the immunodominant epitopes of the AAV-2 capsid surface, it would be possible to rationally design of AAV-2 capsid primarily at those sites.

Rational design involves the specific and purposeful mutation of the viral surface in order to meet a particular need. In this case, rational design would be invoked in

order to create mutate AAV-2 amino acids on the capsid surface in such a way as to alter the capsid to be unrecognizable to antibodies which detect wild-type AAV-2.

Previously, comprehensive mutational studies have been too unwieldy too approach comprehensively due to the sheer number of amino acids on the capsid surface, each of which can be mutated to a possible 20 different amino acids. For CPV, immunodominant sites consisted of 6-8 amino acids. A comprehensive mutation of even these amino acids at 2-3 major immunogenic sites would involve the production and characterization of 240-480 mutants. The most comprehensive mutational analysis accomplished to date on AAV involves 63 single-site mutations and their subsequent characterization (Lochrie et al., 2006).

A more focused approach would involve determining which amino acid mutations are the most important and limiting analyses to those mutations. Most information about the antigenicity of an amino acid can be derived from four mutations including ones based on positive and negative charge, hydrophobicity, and steric properties at the major immunogenic sites. The effect of charge and hydrophobicity could be assessed by mutating an amino acid to an uncharged polar, positively charged, negatively charged, and hydrophobic amino acid of approximately equivalent length. A glutamine, for example, could be mutated to glutamic acid and lysine to determine the effect of replacing an uncharged polar amino acid with a positive or negative charge. The glutamine could then be mutated to leucine to determine the effect of a hydrophobic mutation. Additionally, steric properties can be assessed by mutating an amino acid to one of significantly increased length. For

example, the glutamine could be mutated to tryptophan to determine whether the amino acid can sterically hinder the binding interaction. By characterizing major antigenic sites in this way, the number of mutations is limited to 48-96 for 6-8 amino acids at 2-3 sites. Each mutation can be assessed for escape by challenging the mutant with antibodies that contain the amino acid within their binding footprint.

However, it may be more practical for initial attempts to create escape mutants to focus on even fewer amino acid mutations. The most likely mutations to alter the functionality of a binding site are those which will give the greatest change. For example, the mutation of a positively charged lysine to a negatively charged glutamic acid may significantly alter the binding energy of an antibody bound to that region. Similarly, the relatively large tryptophan or lysine side chains could be mutated into the most exposed amino acid(s) at the epitope to elucidate steric effects. This would limit the number of mutations necessary for initial characterization of an epitope to a relatively small proportion of those previously proposed. Limiting mutational analyses will also allow for the possibility of accelerating mutagenesis assessments against monoclonal (and polyclonal) antibodies prior to a comprehensive understanding of the antigenic surface of AAV-2.

The creation of escape mutants which are useful as gene therapy vectors would rely on the characterizations previously suggested. Escape mutants must be antigenically distinct from the wild-type virus, to which a majority of the U.S.

populace has been previously exposed (Calcedo et al., 2009). It has been shown that a single mutation may significantly inhibit the binding of a monoclonal antibody at a single epitope (Lochrie et al., 2006). However, a single-site escape mutant to a monoclonal antibody is not expected to completely evade neutralization by polyclonal sera. Therefore, an escape mutation must be made at each major antigenic site. The most promising single-site escape mutants may then be combined to create a multi-site mutant that hopefully will yield a new artificial AAV serotype, manufactured to be antigenically distinct from wild-type AAV-2. Creating multiple single-site escape mutants at the same major antigenic sites will lead to the possibility of creating a multiplicity of new AAV serotypes which evade neutralization—as long as the antigenic sites are sufficiently distinct for each new serotype. Escape mutants which are not sufficiently distinct from one another will be neutralized by polyclonal sera to other newly-created serotypes.

We present this rational design approach as a robust means to the creation of a pool of new and antigenically distinct AAV serotypes. Antigentic information learned through this approach would remain applicable even if the virus was retargeted to other cell types. Additionally, antigenic information learned about AAV-2 would be directly transferrable to antigenically similar serotypes, such as AAV-3b. Additionally, we would acquire a broader understanding of whether electrostatic, hydrophobic, or steric mutations are more likely to create viable escape mutations and which topological features of the viral surface are more suitable for creating escape mutants. We believe that these results would further AAV-2's use as a gene

therapy vector and will provide lessons for gene therapy and viral antigenicity as a whole.

APPENDIX I: CREATION OF AAV-2 SINGLE SITE MUTANT PLASMIDS

INTRODUCTION

A major obstacle to the use of AAV-2 as a gene therapy vector is ensuring efficient transduction of the vector in the 35-80% of individuals who have been previously exposed to the virus (Calcedo et al., 2009). In order for AAV to be efficiently transduced in seropositive individuals, new serotypes of AAV which are antigenically distinct from the wild-type must be developed (Agbandje-McKenna and Kleinschmidt, 2011). The rational design of antigenically distinct vectors requires a working knowledge of the vector's major antigenic epitopes. At the outset of this work major epitopes on AAV-2 had only been partially characterized (Girod et al., 1999; Lochrie et al., 2006; Wu et al., 2000), although it is expected that there may be at least two major epitopes on the capsid surface, similar to other members of the family *Parvoviridae* (Hafenstein et al., 2009).

It was for this reason that an in vitro selection approach was pursued by Daniel Mitchell (former student in the laboratory), who attempted to select for escape mutations among naturally occurring variants in AAV-2 by challenging with A20 monclonal antibody (unpublished). The mutation rate for MVM had been reported to be approximately 10^{-5} per nucleotide per replication (Lopez-Bueno, Mateu, and Almendral, 2003). Cellular genes, by contrast, are reported to have a mutation rate of approximately 10^{-9} per nucleotide per replication (Drake, 1991; Goodman and Fygenson, 1998).

After repeated challenge by A20, however, Mitchell only recovered wild-type AAV-2 genetic information from infected cells after sequencing approximately 100 viral clones (Mitchell and Chapman, 2006). These results cast doubt on the reported natural mutation rate of parvoviruses, which may be much lower than initially suggested. It is plausible, therefore, that the actual mutation rate of AAV is less than 10^{-5} for MVM, and potentially as low as the 10^{-9} mutation rate for cellular genes (Greener, Callahan, and Jerpseth, 1996).

The research in this appendix succeeds Daniel Mitchell's work in an attempt to overcome the difficulties faced by the unexpectedly low mutational rate of AAV-2 relative to the reported mutational rate of MVM. The new approach was to randomly mutate the AAV-2 plasmid prior to exerting selective pressure in order to increase the opportunity for natural selection. In the most ideal scenario, a pool of all possible single-site mutations on the plasmid would be created, allowing selection among all possible capsid mutations.

There are several possible ways of introducing random mutations in the AAV-2 plasmid. Error-prone PCR is a popular technique which uses error-prone DNA polymerases to insert as few as 4 DNA mutations per 2000 base pairs, although the mutation frequency may vary depending on the experimental conditions used (Rasila, Pajunen, and Savilahti, 2009). Chemical methods of mutating plasmids have traditionally been accomplished in vivo (Singer and Kusmierek, 1982), although in vitro methods have been successful as well (Lai et al., 2004). Additionally, there are cell-based mutagenesis techniques that use ionizing radiation or ultraviolet light,

which causes genetic mutations indirectly, or cell strains that are deficient in DNA repair pathways.

XL1-Red (Agilent Technologies) is a strain of E. coli which is deficient in three primary DNA repair pathways: error-prone mismatch repair (mutS), 3'- to 5-' exonuclease of DNA polymerase III (mutD), and is unable to hydrolyze 8-oxodGTP (mutT) (Greener, Callahan, and Jerpseth, 1996). The combination of mutations in XL1-Red is reported to give the strain a measured mutation rate 5000-fold higher than wild-type (Greener, Callahan, and Jerpseth, 1996). Therefore, random mutations were introduced into the pAV2 plasmid when propagated in XL1-Red cells (Greener, Callahan, and Jerpseth, 1996). These mutations occurred along the entirety of the pAV2 plasmid, with mutations pertinent to antibody escape occurring in the VP-3 region.

Along the capsid, there are 735 residues, each with a total of 20 possible non-conserved mutations per residue. In the genome, this corresponds to 2205 base pairs, each with a possible 3 mutations. Therefore, a total of 6,615 unique mutations are needed for complete coverage of all possible mutations. Whether or not the mutation is synonymous is based on the individual codon and, on average, 29% of all base pair mutations are synonymous.

METHODS

Creating an AAV-2 Plasmid Pool with Random Single-Site Mutations

XL1-Red competent cells and XL1-Blue competent cells were purchased from Stratagene. XL1-Red cells were used to create single-site mutations and XL1-Blue cells were used to propagate colonies. The heat pulse step has been optimized for 14-mL BD Falcon round-bottomed polypropylene tubes. The Falcon tubes were pre-chilled on ice. 100 µL of XL1-Red cells and 1.7 µL of 1.42 M β-mercaptothanol were added to the chilled tube. The cells were incubated on ice for 10 minutes and were gently swirled every two minutes during the incubation period. 10 ng of pAV2 was then added to the chilled tube and incubated on ice for another 30 minutes. For the heat-pulse step, the cells were added to a 42 °C water bath for 45 seconds and then placed back on ice for 2 minutes. Fresh Super-Optimal broth with Catabolite repression (SOC Media) was pre-warmed to 42 °C. 900 mL of the SOC Media was added to the cells and then incubated at 37 °C for 1 hour on an orbital shaker at 225 rpm.

Selection of XL1-Red cells that have been transformed by the presence of pAV2 was accomplished by first plating 200 µL of the transformation mixture onto a 100 mm lysogeny broth (LB) agar plate with 100mg ampicillin. Colonies which grow on the plates have been transformed with the pAV2 plasmid which confers ampicillin resistance. One colony was then transferred to 15 mL of LB media and incubated for 12 hours at 37 °C. The cells were grown in multiple passages in order to maximize

the number of generations of the transformed XL1-Red cells, and hence maximize the total number of mutations. Therefore, 15 μL of the previous passage was split into 15 mL of LB media, and incubated for 12 hours. This process was repeated a total of 10 times (10 passages). The plasmid was recovered from the 11th passage of XL1-Red cells by performing a miniprep with QIAprep according to the manufacturer's specifications.

The recovered XL1-Red plasmid contained a distribution of genetic mutations. In order to produce genetically identical, cloned plasmids from XL1-Red in sufficient quantity for DNA sequencing it was necessary to grow the plasmid in XL1-Blue. XL1-Blue contains neither mutS, mutD, or mutT and maintains fully functional DNA repair pathways. Thus, plasmid purified from XL1-Red was transformed into XL-1 Blue using the same procedure as before, with the following exceptions: 25 colonies were picked from the LB agar plates with ampicillin and each were each grown individually for one passage. Each of the 25 clones were sent for sequencing by the OHSU MMI Core.

The sequences acquired from the MMI Core were aligned to the native pAV2 gene sequence. Differences between the sequences were identified as single site mutations.

Generational Extension after Plasmid Pool Creation

The methods performed above were independently repeated with two variations. First, an attempt to increase the number and diversity of mutants in the plasmid

pool was approached by increasing the number of starting colonies used to produce the plasmid pool. Therefore, 260 colonies were used in the initiation of the growth of XL1-Red in suspension. Second, the number of passages used to propagate XL1-Red was increased to 20.

RESULTS

Creating an AAV-2 Plasmid Pool with Random Single-Site Mutations

The plasmid sequencing results from the XL1-Red plasmid mutations shows a number of interesting results. Out of 25 colonies that were isolated, 10 showed single-site mutations (Table 5). Six of the mutations at nucleotide position 129 (VP1 amino acid 43) were redundant, which likely result from an early mutation in the pAV2 lineage after transfection into XL1-Red, leaving 5 unique single-site mutations in the sequenced colonies. Additionally, mutations at amino acid 43 and 539 were synonymous, in which mutations at the DNA level were observed while the protein sequence was unchanged. A mutation in colony 14 was also observed, but was outside the open reading frame for VP-1. Sequences for colony 2 and 4 contained non-synonymous mutations in which both the DNA sequence and the predicted protein sequence were altered.

The reported mutation rate of XL1-Red is 1 mutation per 30 generations, per 2 kb (Trower, 2009). The gene size for VP-1 is 2.2 kb. There are approximately 30 generations grown from a single colony to confluence in 15 mL of media (Trower,

2009). We may approximate the total number of generations based on there being approximately 10 generations for each confluent passage split at a 1:1000 ratio and allowed to reach confluence again. We can expect the total number of generations to be equal to the number of generations it takes the first colony to reach confluence in addition to the cumulative total of generations per each passage. The formula is as follows:

$$G = I + nS$$

Where G is the total number of generations, I is the number of generations it takes the first colony to reach confluence, S is the number of generations in each passage, and n is the total number of passages. In our case, for I=30, n=10, and S=10, we observe a total of 130 generations. We can then normalize the reported mutation rate of XL1-Red as 4.77 mutations / 130 generations / 2.2kb. In all, four unique mutations were observed in the open reading frame of VP-1 out of a total of 26 plasmids sequenced, or 15%. Therefore, our observed mutation rate for XL1-Red was 0.15 mutations / 130 generations / 2.2 kb, or 32-fold different from the rate expected by extrapolating the vendor's statistics.

Generational Extension after Plasmid Pool Creation

The total number of generations was extended to 230 using an entirely new sample (the product was not repassaged). Of eight colonies sequenced, four maintained a conserved mutation at position 43, two of which contained an additional L->P mutation at position 647 (Table 6). This leaves two independent mutations

occurring over eight sequenced colonies for the second plasmid pool, equivalent to a 25% mutation rate. The normalized mutation rate is therefore 0.25 mutations / 230 generations / 2.2 kb, or 19-fold different from the rate expected by extrapolating the vendor's statistics.

Colony	Genetic Mutation	Amino Acid Mutation	VP1 Amino Acid
2	G -> A	A -> T	655
4	G -> A	G -> E	111
5,7,8,17,20,25	G -> A	Synonymous	43
9	T -> C	Synonymous	539
14	G -> A	N/A	Outside capsid region

Table 5. Single-site mutations of pAV2 plasmid after proliferation in XL1-Red. Plasmid pool was derived from a single colony and grown for 10 passages.

Colony	Genetic Mutation	Amino Acid Mutation	VP1 Amino Acid
2,3,7,8	G->A	Synonymous	43
2,3	T->C	L->P	647

Table 6. Single-site mutations of pAV2 plasmid after proliferation in XL1-Red. Plasmid pool was derived from multiple colonies and grown for 20 passages.

DISCUSSION

The original objective of creating XL1-Red plasmid mutants was to create a plasmid pool of all possible single-site capsid mutations and select for escape mutants. In order to accomplish this objective it would be necessary to create a plasmid pool which contains 6,615 unique plasmids. In our two limited tests, we observed 5 plasmid mutants in one and 2 plasmid mutants in the other.

Surprisingly, one of mutation occurred in both the 130 generation library and the 230 generation library. With the current data sets, the probability of a mutation being randomly generated in both sample sets is greater than 3 standard deviations away from the mean. Therefore, it is very unlikely that the mutation occurred in both sample sets randomly. It is a possibility that the original plasmid stock already contained the mutation. In the absence of this scenario, it is likely that the XL1-Red does not mutate amino acid positions entirely at random, but rather preferentially mutates certain amino acid positions.

An additional point of interest relative to the randomness of XL1-Red mutations resides in which base-pairs are mutated to which new base-pair. In a hypothetical situation where randomly distributed DNA was mutated at random, we would expect to observe G, A, C, and T each to be mutated 25% of the time over a large data set. Additionally, each specific G -> A, G -> C, and G -> T mutation would be expected to occur 8.3% of the time. In this case, a total of 7 mutations were generated. Five were G -> A mutations and two were T -> C mutations. Chi-squared test results suggest that it is extremely statistically significant that the distribution of mutations is not the hypothesized random mutation distribution.

In retrospect, it is perhaps expected that XL1-Red preferentially induced transition mutations (A↔G or C↔T). A recent evaluation of XL1-Red and other random mutagenesis protocols reported that XL1-Red transition mutations occurred six times more frequently than transversion mutations (A↔C or G↔T), although this is only over a sample size of 7 mutations (Rasila, Pajunen, and Savilahti, 2009). In our

case, mutations are exclusively transitions, suggesting that transition mutations may occur more frequently than reported by Rasila et al.

While the XL1-Red is not a perfect system, it may be sufficient for sampling the complete array of all possible mutations. AAV-2 contains 736 capsid amino acids. We have observed a unique mutation rate of 25%. Non-synonymous mutations are, on average, 71% of all possible random mutations at the genetic level. The transduction efficiency of AAV-2 is approximately 1/2000 in HeLa cells (Grimm et al., 2008). Therefore, approximately 8.3×10^6 plasmids would need to be acquired in order to transfect HeLa cells with AAV-2 plasmids containing one mutation per capsid amino acid, when transversion preferences are not considered.

The volume of bacteria necessary to create this many plasmids may be similarly calculated. The copy number of pBluescript is 400. There are approximately 1×10^9 E. coli bacteria / mL near confluence. Therefore we expect that 0.021 microliters of bacteria will contain enough plasmids to sample one mutation per capsid amino acid.

The mass of plasmid necessary to create the requisite number of plasmids may also be calculated. pBluescript is a 3.0 kb phagmid vector that contains the 5.0 kb AAV-2 AAV-2 sequence. On average, one base pair is approximately 660 kDa, or 5,280 MDa for the 8 kb plasmid. One gram is equivalent to 6.022×10^{23} Da. Therefore, it would require 73 ng of plasmid for the sample to contain enough plasmids to sample one mutation per capsid amino acid. We expect this mass of plasmid to be feasible to use in selection experiments.

We expect XL1-Red to be a viable host for acquiring a large population of AAV single-site mutants. The XL1-Red system is rapid and reproducible, however, a large population of mutants could also be acquired though sister-methods such as error-prone PCR. Error-prone PCR also has advantages in that a specific region within the gene can be targeted, the mutational frequency may be optimized at a higher frequency than XL1-Red, and transversion mutations occur significantly more frequently than in an XL1-Red system (Rasila, Pajunen, and Savilahti, 2009). In this regard, the benefits of error-prone PCR could overcome some difficulties encountered when using XL1-Red to create random mutations. However, error-prone PCR also has several disadvantages. One drawback of error-prone PCR is that it must be optimized for the mutational frequency desired (Greener, Callahan, and Jerpseth, 1996). Additionally, the current lower limit on error-prone PCR mutagenesis corresponds to an average of 3 coding changes per capsid protein (Rasila, Pajunen, and Savilahti, 2009), which leaves ambiguity in the correlation of phenotype to genotype. Until such time that this can be reduced another 3-fold, the XL1-Red system, with its lower mutation rate, offers a more direct analysis of the roles of individual amino acids.

REFERENCES

Agbandje-McKenna, M., and Chapman, M. S. (2006). Correlating structure with function in the viral capsid. *In* "Parvoviruses" (J. R. Kerr, S. F. Cotmore, M. E. Bloom, R. M. Linden, and C. R. Parrish, Eds.), pp. 124-139. Hodder Arnold, Ltd., London.

Agbandje-McKenna, M., and Kleinschmidt, J. (2011). AAV capsid structure and cell interactions. *Methods Mol Biol* **807,** 47-92.

Alam, S. M., McAdams, M., Boren, D., Rak, M., Scearce, R. M., Gao, F., Camacho, Z. T., Gewirth, D., Kelsoe, G., Chen, P., and Haynes, B. F. (2007). The role of antibody polyspecificity and lipid reactivity in binding of broadly neutralizing anti-HIV-1 envelope human monoclonal antibodies 2F5 and 4E10 to glycoprotein 41 membrane proximal envelope epitopes. *J Immunol* **178**(7)**,** 4424-35.

Atchison, R. W., Casto, B. C., and Hammon, W. M. (1965). Adenovirus-Associated Defective Virus Particles. *Science* **149,** 754-6.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**(18)**,** 10037-41.

Balague C, K. M., Zhang WW (1997). Adeno-associated virus Rep78 protein and terminal repeats enhance intergration of DNA sequences into the cellular genome. *J Virol* **71**(4)**,** 3299-3306.

Bartlett, J. S., Wilcher, R., and Samulski, R. J. (2000). Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors. *J Virol* **74**(6)**,** 2777-85.

Berns, K. I., and Labow, M. A. (1987). Parvovirus gene regulation. *J Gen Virol* **68 (Pt 3),** 601-14.

Berns, K. I., and Linden, R. M. (1995). The cryptic life style of adeno-associated virus. *Bioessays* **17**(3)**,** 237-45.

Blacklow, N. R., Hoggan, M. D., and Rowe, W. P. (1968). Serologic evidence for human infection with adenovirus-associated viruses. *J Natl Cancer Inst* **40**(2)**,** 319-27.

Bleker, S., Sonntag, F., and Kleinschmidt, J. A. (2005). 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. *Journal of Virology* **79**(4)**,** 2528-2540.

Boutin, S., Monteilhet, V., Veron, P., Leborgne, C., Benveniste, O., Montus, M. F., and Masurier, C. (2010). 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* **21**(6)**,** 704-12.

Brockstedt, D. G., Podsakoff, G. M., Fong, L., Kurtzman, G., Mueller-Ruchholtz, W., and Engleman, E. G. (1999). Induction of immunity to antigens expressed by recombinant adeno-associated virus depends on the route of administration. *Clin Immunol* **92**(1)**,** 67-75.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998). Crystallography and NMR system: A new software system for macromolecular structure determination. *Acta Crystallographica* **D54,** 905-921.

Bueler, H. (1999). Adeno-associated viral vectors for gene transfer and gene therapy. *Biol Chem* **380**(6)**,** 613-22.

Calcedo, R., Vandenberghe, L. H., Gao, G., Lin, J., and Wilson, J. M. (2009). Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses. *J Infect Dis* **199**(3)**,** 381-90.

Carter, B., and Flotte, T. (1996). Development of adeno-associated virus vectors for gene therapy of cystic fibrosis. *Current Topics in Microbiology and Immunology* **218,** 119-144.

Carter, B. J. (2005). Adeno-associated virus vectors in clinical trials. *Hum Gene Ther* **16**(5)**,** 541-50.

Carter, B. J., Burstein, H., and Peluso, R. W. (2008). Adeno-associated Virus and AAV Vectors for Gene delivery. *In* "Gene and cell therapy: therapeutic mechanisms and strategies" (N. S. Templeton, Ed.), pp. 115-156. CRC Press, Boca Raton.

Caspar, D. L. D., and Klug, A. (1962). Physical principles in the constructuion of regular viruses. *Cold Spring Harbor Symposium in Quantitative Biology* **27,** 1-24.

Chapman, M. S. (1995). Restrained Real-Space Macromolecular Atomic Refinement using a New Resolution-Dependent Electron Density Function. *Acta Crystallographica* **A51,** 69-80.

Chapman, M. S., and Agbandje-McKenna, M. (2006). Atomic structure of viral particles. *In* "Parvoviruses" (J. R. Kerr, S. F. Cotmore, M. E. Bloom, R. M. Linden, and C. R. Parrish, Eds.), pp. 107-123. Hodder Arnold, Ltd., London.

Chejanovsky, N., and Carter, B. J. (1989). Mutagenesis of an AUG codon in the adeno-associated virus rep gene: effects on viral DNA replication. *Virology* **173**(1)**,** 120-8.

Chirmule, N., Propert, K., Magosin, S., Qian, Y., Qian, R., and Wilson, J. (1999). Immune responses to adenovirus and adeno-associated virus in humans. *Gene Ther* **6**(9)**,** 1574-1583.

Chirmule, N., Xiao, W., Truneh, A., Schnell, M. A., Hughes, J. V., Zoltick, P., and Wilson, J. M. (2000). Humoral immunity to adeno-associated virus type 2 vectors following administration to murine and nonhuman primate muscle. *J Virol* **74**(5)**,** 2420-5.

Cohen, S., Behzad, A. R., Carroll, J. B., and Pante, N. (2006). Parvoviral nuclear import: bypassing the host nuclear-transport machinery. *J Gen Virol* **87**(Pt 11)**,** 3209-13.

Cohen, S., and Pante, N. (2005). Pushing the envelope: microinjection of Minute virus of mice into Xenopus oocytes causes damage to the nuclear envelope. *J Gen Virol* **86**(Pt 12)**,** 3243-52.

Cole, A. (2008). Child in gene therapy programme develops leukaemia. *BMJ* **336**(7634)**,** 13.

Cossart, Y. E., Field, A. M., Cant, B., and Widdows, D. (1975). Parvovirus-like particles in human sera. *Lancet* **1**(7898)**,** 72-3.

Davidoff, A. M., Gray, J. T., Ng, C. Y., Zhang, Y., Zhou, J., Spence, Y., Bakar, Y., and Nathwani, A. C. (2005). Comparison of the ability of adeno-associated viral vectors pseudotyped with serotype 2, 5, and 8 capsid proteins to mediate efficient transduction of the liver in murine and nonhuman primate models. *Mol Ther* **11**(6)**,** 875-88.

Dong, J. Y., Fan, P. D., and Frizzell, R. A. (1996). Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. *Human Gene Therapy* **7**(17)**,** 2101-2112.

Douar, A. M., Poulard, K., Stockholm, D., and Danos, O. (2001). Intracellular trafficking of adeno-associated virus vectors: routing to the late endosomal compartment and proteasome degradation. *J Virol* **75**(4)**,** 1824-33.

Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A* **88**(16)**,** 7160-4.

Dubielzig, R., King, J. A., Weger, S., Kern, A., and Kleinschmidt, J. A. (1999). Adeno-associated virus type 2 protein interactions: formation of pre-encapsidation complexes. *J Virol* **73**(11)**,** 8989-98.

Dunbrack Jr., R. L., and Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science* **1661-81**.

Dunbrack, R. L., Jr. (2002). Rotamer libraries in the 21st century. *Curr Opin Struct Biol* **12**(4)**,** 431-40.

Dutheil, N., Shi, F., Dupressoir, T., and Linden, R. M. (2000). Adeno-associated virus site-specifically integrates into a muscle-specific DNA region. *Proc Natl Acad Sci U S A* **97**(9)**,** 4862-6.

Eisenberg, D., and McLachlan, A. D. (1986). Solvation energy in protein folding and binding. *Nature* **319,** 199-203.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 4)**,** 486-501.

Eswar, N., Eramian, D., Webb, B., Shen, M., and Sali, A. (2006). Protein Structure Modeling With MODELLER. *In* "Current Protocols in Bioinformatics" (A. D. Baxevanis, L. D. Stein, G. D. Stormo, and J. R. Yates III, Eds.), Vol. Supplement 15, pp. 5.6.1-5.6.30. John Wiley & Sons, Inc.

Fabiola, F., and Chapman, M. S. (2005). Fitting of high-resolution structures into electron microscopy reconstruction images. *Structure (Camb)* **13**(3)**,** 389-400.

Farr, G. A., Zhang, L. G., and Tattersall, P. (2005). Parvoviral virions deploy a capsid-tethered lipolytic enzyme to breach the endosomal membrane during cell entry. *Proc Natl Acad Sci U S A* **102**(47)**,** 17148-53.

Fauchere, J.-L., and Pliska, V. (1983). Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. med. Chem.-Chim. Ther.* **18,** 369-75.

Fernandez, J. J., Luque, D., Caston, J. R., and Carrascosa, J. L. (2008). Sharpening high resolution information in single particle electron cryomicroscopy. *J Struct Biol* **164**(1)**,** 170-5.

Ferrari FK, S. T., Shenk T, Samulski RJ (1996). Second-strand synthesis is a rate-limiting step for sfficient transduction by recombinant adeno-associated virus vectors. *J Virol* **70**(5)**,** 3227-3234.

Fisher KJ, G. G., Weitzman MD, DeMatteo R, Burda JF, Wilson JM (1996). Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. *J Virol* **70**(1)**,** 520-532.

Flotte, T. R. (2005). Recent developments in recombinant AAV-mediated gene therapy for lung diseases. *Curr Gene Ther* **5**(3)**,** 361-6.

Flotte, T. R. (2007). Gene therapy: the first two decades and the current state-of-the-art. *J Cell Physiol* **213**(2)**,** 301-5.

Flotte, T. R., and Carter, B. J. (1998). Adeno-associated virus vectors for gene therapy of cystic fibrosis. *Methods Enzymol* **292,** 717-32.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**(1)**,** 190-9.

Gao, G., Vandenberghe, L. H., Alvira, M. R., Lu, Y., Calcedo, R., Zhou, X., and Wilson, J. M. (2004). Clades of Adeno-associated viruses are widely disseminated in human tissues. *J Virol* **78**(12)**,** 6381-8.

Gao, G., Vandenberghe, L. H., and Wilson, J. M. (2005). New recombinant serotypes of AAV vectors. *Curr Gene Ther* **5**(3)**,** 285-97.

Gerstein, M., and Levitt, M. (1998). Comprehensive assessment of automatic structural alignment against a manual standard, the scop classification of proteins. *Protein Sci* **7**(2)**,** 445-56.

Gerstein, M., Richards, F., Chapman, M. S., and Connolly, M. (2001). Protein surfaces and volumes: measurement and use. *In* "International Tables for Crystallography. Crystallography of Biological Molecules." (M. G. Rossmann, and E. Arnold, Eds.), Vol. F, pp. 531-45 (Cpt. 22.1). Kluwer Academic Publishers, Dortrecht, Netherlands.

Girod, A., Ried, M., Wobus, C., Lahm, H., Leike, K., Kleinschmidt, J., Deleage, G., and Hallek, M. (1999). Genetic capsid modifications allow efficient re-targeting of adeno- associated virus type 2. *Nat Med* **5**(9)**,** 1052-1056.

Girod, A., Wobus, C. E., Zadori, Z., Ried, M., Leike, K., Tijssen, P., Kleinschmidt, J. A., and Hallek, M. (2002). The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. *J Gen Virol* **83**(Pt 5)**,** 973-8.

Glas, C. A., and van der Linden, W. J. (2010). Marginal likelihood inference for a model for item responses and response times. *Br J Math Stat Psychol* **63**(Pt 3)**,** 603-26.

Goodman, M. F., and Fygenson, K. D. (1998). DNA polymerase fidelity: from genetics toward a biochemical understanding. *Genetics* **148**(4)**,** 1475-82.

Govindasamy, L., Padron, E., McKenna, R., Muzyczka, N., Kaludov, N., Chiorini, J. A., and Agbandje-McKenna, M. (2006). Structurally mapping the diverse phenotype of adeno-associated virus serotype 4. *J Virol* **80**(23)**,** 11556-70.

Greener, A., Callahan, M., and Jerpseth, B. (1996). An efficient random mutagenesis technique using an E. coli mutator strain. *Methods Mol Biol* **57,** 375-85.

Grimm, D., Lee, J. S., Wang, L., Desai, T., Akache, B., Storm, T. A., and Kay, M. A. (2008). In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J Virol* **82**(12)**,** 5887-911.

Hafenstein, S., Bowman, V. D., Sun, T., Nelson, C. D., Palermo, L. M., Chipman, P. R., Battisti, A. J., Parrish, C. R., and Rossmann, M. G. (2009). Structural comparison of different antibodies interacting with parvovirus capsids. *J Virol*.

Hafenstein, S., Palermo, L. M., Kostyuchenko, V. A., Xiao, C., Morais, M. C., Nelson, C. D., Bowman, V. D., Battisti, A. J., Chipman, P. R., Parrish, C. R., and Rossmann, M. G. (2007). Asymmetric binding of transferrin receptor to parvovirus capsids. *Proc Natl Acad Sci U S A* **104**(16)**,** 6585-9.

Halbert, C. L., Rutledge, E. A., Allen, J. M., Russell, D. W., and Miller, A. D. (2000). Repeat transduction in the mouse lung by using adeno-associated virus vectors with different serotypes. *J Virol* **74**(3)**,** 1524-32.

Halbert, C. L., Standaert, T. A., Aitken, M. L., Alexander, I. E., Russell, D. W., and Miller, A. D. (1997). Transduction by adeno-associated virus vectors in the rabbit airway: efficiency, persistence, and readministration. *J Virol* **71**(8)**,** 5932-41.

Hansen, J., Qing, K., and Srivastava, A. (2001). Adeno-associated virus type 2-mediated gene transfer: altered endocytic processing enhances transduction efficiency in murine fibroblasts. *J Virol* **75**(9)**,** 4080-90.

Hauck, B., and Xiao, W. (2003). Characterization of tissue tropism determinants of adeno-associated virus type 1. *J Virol* **77**(4)**,** 2768-74.

Hernandez, Y. J., Wang, J., Kearns, W. G., Loiler, S., Poirier, A., and Flotte, T. R. (1999). Latent adeno-associated virus infection elicits humoral but not cell-mediated immune responses in a nonhuman primate model. *J Virol* **73**(10)**,** 8549-58.

Hewat, E. A., and Blaas, D. (1996). Structure of a neutralizing antibody bound bivalently to human rhinovirus 2. *EMBO J* **15**(7)**,** 1515-23.

Hewat, E. A., Marlovits, T. C., and Blaas, D. (1998). Structure of a neutralizing antibody bound monovalently to human rhinovirus 2. *J Virol* **72**(5)**,** 4396-402.

Hirosue, S., Senn, K., Clement, N., Nonnenmacher, M., Gigout, L., Linden, R. M., and Weber, T. (2007). Effect of inhibition of dynein function and microtubule-altering drugs on AAV2 transduction. *Virology* **367**(1)**,** 10-8.

Hodgson, C. P. (1995). The Vector Void in Gene Therapy. *Bio/Technology* **13,** 222-225.

Hoque, M., Ishizu, K., Matsumoto, A., Han, S. I., Arisaka, F., Takayama, M., Suzuki, K., Kato, K., Kanda, T., Watanabe, H., and Handa, H. (1999). Nuclear transport of the major capsid protein is essential for adeno- associated virus capsid formation. *J Virol* **73**(9)**,** 7912-5.

Im, D.-S., and Muzyczka, N. (1990). The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. *Cell* **61,** 447-457.

Ito, M., Tchoua, U., Okamoto, M., and Tojo, H. (2002). Purification and properties of a phospholipase A2/lipase preferring phosphatidic acid, bis(monoacylglycerol) phosphate, and monoacylglycerol from rat testis. *J Biol Chem* **277**(46)**,** 43674-81.

Izui, S., Berney, T., Shibata, T., and Fulpius, T. (1993). IgG3 cryoglobulins in autoimmune MRL-lpr/lpr mice: immunopathogenesis, therapeutic approaches and relevance to similar human diseases. *Ann Rheum Dis* **52 Suppl 1,** S48-54.

Kaludov, N., Brown, K. E., Walters, R. W., Zabner, J., and Chiorini, J. A. (2001). 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* **75**(15)**,** 6884-93.

Kaufmann, B., Lopez-Bueno, A., Mateu, M. G., Chipman, P. R., Nelson, C. D., Parrish, C. R., Almendral, J. M., and Rossmann, M. G. (2007). Minute virus of mice, a parvovirus, in complex with the Fab fragment of a neutralizing monoclonal antibody. *J Virol* **81**(18)**,** 9851-8.

Kay, M. A., Manno, C. S., Ragni, M. V., Larson, P. J., Couto, L. B., McClelland, A., Glader, B., Chew, A. J., Tai, S. J., Herzog, R. W., Arruda, V., Johnson, F., Scallan, C., Skarsgard, E., Flake, A. W., and High, K. A. (2000). Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector. *Nat Genet* **24**(3)**,** 257-61.

Kelkar, S., De, B. P., Gao, G., Wilson, J. M., Crystal, R. G., and Leopold, P. L. (2006). A common mechanism for cytoplasmic dynein-dependent microtubule binding shared among adeno-associated virus and adenovirus serotypes. *J Virol* **80**(15)**,** 7781-5.

Kerr, J. R. (2006). "Parvoviruses." Hodder Arnold ;

Distributed in the United States of America by Oxford University Press, London

New York.

King, A. (2012). "Virus taxonomy : classification and nomenclature of viruses : ninth report of the International Committee on Taxonomy of Viruses." Academic Press, London ; Waltham, MA.

King, J. A., Dubielzig, R., Grimm, D., and Kleinschmidt, J. A. (2001). DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *Embo J* **20**(12)**,** 3282-91.

Kotin, R. M., Linden, R. M., and Berns, K. I. (1992). Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J* **11**(13)**,** 5071-8.

Kotin, R. M., Siniscalco, M., Samulski, R. J., Zhu, X. D., Hunter, L., Laughlin, C. A., McLaughlin, S., Muzyczka, M., Rocchi, M., and Berns, K. I. (1990). Site-specific integration by adeno-associated virus. *Proceedings of the National Academy of Sciences, USA* **87,** 2211-5.

Kronenberg, S., Bottcher, B., von der Lieth, C. W., Bleker, S., and Kleinschmidt, J. A. (2005). A conformational change in the adeno-associated virus type 2 capsid leads to the exposure of hidden VP1 N termini. *J Virol* **79**(9)**,** 5296-303.

Kronenberg, S., Kleinschmidt, J. A., and Bottcher, B. (2001). Electron cryo-microscopy and image reconstruction of adeno-associated virus type 2 empty capsids. *EMBO Rep* **2**(11)**,** 997-1002.

Lai, Y. P., Huang, J., Wang, L. F., Li, J., and Wu, Z. R. (2004). A new approach to random mutagenesis in vitro. *Biotechnol Bioeng* **86**(6)**,** 622-7.

Lander, G. C., Stagg, S. M., Voss, N. R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P. W., Lyumkis, D., Potter, C. S., and Carragher, B. (2009). Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J Struct Biol* **166**(1)**,** 95-102.

Laughlin, C. A., Tratschin, J. D., Coon, H., and Carter, B. J. (1983). Cloning of infectious adeno-associated virus genomes in bacterial plasmids. *Gene* **23**(1)**,** 65-73.

Li, H., Dunn, J. J., Luft, B. J., and Lawson, C. L. (1997). Crystal structure of Lyme disease antigen outer surface protein A complexed with an Fab. *Proc Natl Acad Sci U S A* **94**(8)**,** 3584-9.

Lochrie, M. A., Tatsuno, G. P., Christie, B., McDonnell, J. W., Zhou, S., Surosky, R., Pierce, G. F., and Colosi, P. (2006). Mutations on the external surfaces of

adeno-associated virus type 2 capsids that affect transduction and neutralization. *J Virol* **80**(2)**,** 821-34.

Lopez-Bueno, A., Mateu, M. G., and Almendral, J. M. (2003). High mutant frequency in populations of a DNA virus allows evasion from antibody therapy in an immunodeficient host. *J Virol* **77**(4)**,** 2701-8.

Lovell, S. C., Word, J. M., Richardson, J. S., and Richardson, D. C. (2000). The penultimate rotamer library. *Proteins* **40**(3)**,** 389-408.

Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**(1)**,** 82-97.

Luo, M., Tsao, J., Rossmann, M. G., Bassak, S., and Compans, R. W. (1988). Preliminary X-ray crystallographic analysis of Canine Parvovirus Crystals. *J. Mol. Biol.* **200,** 209-211.

Mallick, S. P., Carragher, B., Potter, C. S., and Kriegman, D. J. (2005). ACE: automated CTF estimation. *Ultramicroscopy* **104**(1)**,** 8-29.

Manning, W. C., Zhou, S., Bland, M. P., Escobedo, J. A., and Dwarki, V. (1998). Transient immunosuppression allows transgene expression following readministration of adeno-associated viral vectors. *Hum Gene Ther* **9**(4)**,** 477-85.

Manno, C. S., Pierce, G. F., Arruda, V. R., Glader, B., Ragni, M., Rasko, J. J., Ozelo, M. C., Hoots, K., Blatt, P., Konkle, B., Dake, M., Kaye, R., Razavi, M., Zajko, A., Zehnder, J., Rustagi, P. K., Nakai, H., Chew, A., Leonard, D., Wright, J. F., Lessard, R. R., Sommer, J. M., Tigges, M., Sabatino, D., Luk, A., Jiang, H., Mingozzi, F., Couto, L., Ertl, H. C., High, K. A., and Kay, M. A. (2006). Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat Med* **12**(3)**,** 342-7.

Marshall, E. (1999). Gene Therapy Vector Death Prompts Review of Adenovirus Vector. *Science* **286,** 2244-5.

Mingozzi, F., and High, K. A. (2011). Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. *Nat Rev Genet* **12**(5)**,** 341-55.

Mitchell, D. (2008). Florida State University, Tallahassee.

Mitchell, D. A. J., and Chapman, M. S. (2006). Adeno-Associated Virus (AAV) Antibody Neutralization Escape Mutants: Toward the Characterization of Viral Epitopes. *In* "American Society for Virology Annual Meeting", Madison, WI.

Moskalenko, M., Chen, L., van Roey, M., Donahue, B. A., Snyder, R. O., McArthur, J. G., and Patel, S. D. (2000). Epitope mapping of human anti-adeno-associated virus type 2 neutralizing antibodies: implications for gene therapy and virus structure. *J Virol* **74**(4)**,** 1761-6.

Mouw, M. B., and Pintel, D. J. (2000). Adeno-associated virus RNAs appear in a temporal order and their splicing is stimulated during coinfection with adenovirus. *J Virol* **74**(21)**,** 9878-88.

Murphy, S. L., Bhagwat, A., Edmonson, S., Zhou, S., and High, K. A. (2008). High-throughput Screening and Biophysical Interrogation of Hepatotropic AAV. *Mol Ther* **16**(12)**,** 1960-1967.

Muzyczka, N., and Berns, K. I. (2001). *Parvoviridae*: The Viruses and Their Replication. 4 ed. *In* "Virology" (B. N. Fields, D. M. Knipe, and P. M. Howley, Eds.), pp. 2327-2360. 2 vols. Lippincott Williams & Wilkins, Philadelphia.

Myers, M. W., and Carter, B. J. (1980a). Assembly of adeno-associated virus. *Virology* **102**(1)**,** 71-82.

Myers, M. W., and Carter, B. J. (1980b). Assembly of adeno-associated virus. *Virology* **102,** 71-82.

Naldini, L. (2011). Ex vivo gene transfer and correction for cell-based therapies. *Nat Rev Genet* **12**(5)**,** 301-15.

Nash, K., Chen, W., and Muzyczka, N. (2008). Complete in vitro reconstitution of adeno-associated virus DNA replication requires the minichromosome maintenance complex proteins. *J Virol* **82**(3)**,** 1458-64.

Ni, T. H., Zhou, X., McCarty, D. M., Zolotukhin, I., and Muzyczka, N. (1994). In vitro replication of adeno-associated virus DNA. *Journal of Virology* **68,** 1128-38.

North, B., Lehmann, A., and Dunbrack, R. L., Jr. (2011). A new clustering of antibody CDR loop conformations. *J Mol Biol* **406**(2)**,** 228-56.

O'Donnell, J., Chapman, M. S., and Taylor, K. A. (2008). *Three Dimensional Electron Microscopy, Lucca, Italy*.

O'Donnell, J., Taylor, K. A., and Chapman, M. S. (2009). Adeno-associated virus-2 and its primary cellular receptor-Cryo-EM structure of a heparin complex. *Virology* **385,** 434-443.

Owens, R. A., Weitzman, M. D., Kyostio, S. R., and Carter, B. J. (1993). Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. *J Virol* **67**(2)**,** 997-1005.

Peden, C. S., Burger, C., Muzyczka, N., and Mandel, R. J. (2004). Circulating anti-wild-type adeno-associated virus type 2 (AAV2) antibodies inhibit recombinant AAV2 (rAAV2)-mediated, but not rAAV5-mediated, gene transfer in the brain. *J Virol* **78**(12)**,** 6344-59.

Pereira, D. J., McCarty, D. M., and Muzyczka, N. (1997). The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. *J Virol* **71**(2)**,** 1079-88.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**(13)**,** 1605-12.

Pfeifer, A., and Verma, I. (2001). Virus Vectors and Their Applications. 4 ed. *In* "Virology" (B. N. Fields, D. M. Knipe, and P. M. Howley, Eds.), pp. 469-492. 2 vols. Lippincott Williams & Wilkins, Philadelphia.

Qing, K., Wang, X. S., Kube, D. M., Ponnazhagan, S., Bajpai, A., and Srivastava, A. (1997). Role of tyrosine phosphorylation of a cellular protein in adeno-associated virus 2-mediated transgene expression. *Proc Natl Acad Sci U S A* **94**(20)**,** 10879-84.

Rasila, T. S., Pajunen, M. I., and Savilahti, H. (2009). Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment. *Anal Biochem* **388**(1)**,** 71-80.

Rhode, S. L., 3rd (1973). Replication process of the parvovirus H-1. I. Kinetics in a parasynchronous cell system. *J Virol* **11**(6)**,** 856-61.

Rivadeneira, E. (1998). Sites of recombiant adeno-associated virus integration. *International Journal of Oncology* **12**(4)**,** 805-810.

Rossmann, M. G. (2000). Fitting atomic models into electron-microscopy maps. *Acta Crystallogr D Biol Crystallogr* **56**(Pt 10)**,** 1341-1349.

Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H., Johnson, J. E., Kamer, G., Luo, M., Mosser, A., Rueckert, R., Sherry, B., and Vriend, G. (1985). Structure of human common cold virus and functional relationship to other picornaviruses. . *Nature* **317** 145-153

Rutledge, E., Halbert, C., and Russell, D. (1998). Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. *Journal of Virology* **72**(1)**,** 309-319.

Samulski, R. J. (2000). Expanding the AAV package. *Nat Biotechnol* **18**(5)**,** 497-8.

Samulski, R. J., Sally, M., and Muzyczka, N. (1999). Adeno-associated viral vectors. *In* "The development of human gene therapy" (T. Friedman, Ed.), pp. 132-172. Cold Spring Harbor Laboartory Press, Cold Spring Harbor.

Samulski, R. J., Zhu, X., Xiao, X., Brook, J. D., Housman, D. E., Epstein, N., and Hunter, L. A. (1991). Targeted Integration of Adenoassociated virus (AAV) into Human Cheomosome-19. *EMBO J.* **10,** 3941-50.

Sanlioglu, S., Benson, P. K., Yang, J., Atkinson, E. M., Reynolds, T., and Engelhardt, J. F. (2000). Endocytosis and nuclear trafficking of adeno-associated virus type 2 are controlled by rac1 and phosphatidylinositol-3 kinase activation. *J Virol* **74**(19)**,** 9184-96.

Schmidt, M., Voutetakis, A., Afione, S., Zheng, C., Mandikian, D., and Chiorini, J. A. (2008). Adeno-associated virus type 12 (AAV12): a novel AAV serotype with sialic acid- and heparan sulfate proteoglycan-independent transduction activity. *J Virol* **82**(3)**,** 1399-406.

Schwartz, R. A., Palacios, J. A., Cassell, G. D., Adam, S., Giacca, M., and Weitzman, M. D. (2007). The Mre11/Rad50/Nbs1 complex limits adeno-associated virus transduction and replication. *J Virol* **81**(23)**,** 12936-45.

Seisenberger, G., Ried, M. U., Endress, T., Buning, H., Hallek, M., and Brauchle, C. (2001). Real-time single-molecule imaging of the infection pathway of an adeno-associated virus. *Science* **294**(5548)**,** 1929-32.

Shi, W., Arnold, G. S., and Bartlett, J. S. (2001). 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* **12**(14)**,** 1697-711.

Shi, Y., Seto, E., Chang, L. S., and Shenk, T. (1991). Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* **67**(2)**,** 377-88.

Shneerson, J. M., Mortimer, P. P., and Vandervelde, E. M. (1980). Febrile illness due to a parvovirus. *Br Med J* **280**(6231)**,** 1580.

Singer, B., and Kusmierek, J. T. (1982). Chemical mutagenesis. *Annu Rev Biochem* **51,** 655-93.

Slanina, H., Weger, S., Stow, N. D., Kuhrs, A., and Heilbronn, R. (2006). Role of the herpes simplex virus helicase-primase complex during adeno-associated virus DNA replication. *J Virol* **80**(11)**,** 5241-50.

Smith, T. J., Chase, E. S., Schmidt, T. J., Olson, N. H., and Baker, T. S. (1996). Neutralizing antibody to human rhinovirus 14 penetrates the receptor-binding canyon. *Nature* **383**(6598)**,** 350-4.

Smith, T. J., Olson, N. H., Cheng, R. H., Liu, H., Chase, E. S., Lee, W. M., Leippe, D. M., Mosser, A. G., Rueckert, R. R., and Baker, T. S. (1993). Structure of human rhinovirus complexed with Fab fragments from a neutralizing antibody. *J Virol* **67**(3)**,** 1148-58.

Sonntag, F., Schmidt, K., and Kleinschmidt, J. A. (2010). A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proc Natl Acad Sci U S A* **107**(22)**,** 10220-5.

Stahnke, S., Lux, K., Uhrig, S., Kreppel, F., Hosel, M., Coutelle, O., Ogris, M., Hallek, M., and Buning, H. (2011). Intrinsic phospholipase A2 activity of adeno-associated virus is involved in endosomal escape of incoming particles. *Virology* **409**(1)**,** 77-83.

Suikkanen, S., Antila, M., Jaatinen, A., Vihinen-Ranta, M., and Vuento, M. (2003). Release of canine parvovirus from endocytic vesicles. *Virology* **316**(2)**,** 267-80.

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C. S., and Carragher, B. (2005). Automated molecular microscopy: the new Leginon system. *J Struct Biol* **151**(1)**,** 41-60.

Summerford, C., and Samulski, R. J. (1998). Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J Virol* **72**(2)**,** 1438-45.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**(1)**,** 38-46.

Tattersall, P. (1972). Replication of the parvovirus MVM. I. Dependence of virus multiplication and plaque formation on cell growth. *J Virol* **10**(4)**,** 586-90.

Thomas, C. E., Storm, T. A., Huang, Z., and Kay, M. A. (2004). Rapid uncoating of vector genomes is the key to efficient liver transduction with pseudotyped adeno-associated virus vectors. *J Virol* **78**(6)**,** 3110-22.

Thouvenin, E., Laurent, S., Madelaine, M. F., Rasschaert, D., Vautherot, J. F., and Hewat, E. A. (1997). Bivalent binding of a neutralising antibody to a calicivirus involves the torsional flexibility of the antibody hinge. *J Mol Biol* **270**(2)**,** 238-46.

Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008). Protein Structure Fitting and Refinement Guided by Cryo-EM Density. *Structure* **16**(2)**,** 295-307.

Trower, M. K., Ed. (2009). Methods in Molecular Biology. Vol. 57. Totowa, NJ: Humana Press Inc.

Urcelay, E., Ward, P., Wiener, S. M., Safer, B., and Kotin, R. M. (1995). Asymmetric replication in vitro from a human sequence element is dependent on adeno-associated virus Rep protein. *J Virol* **69**(4)**,** 2038-46.

Van Regenmortel, M. H. V. (1992). "Structure of antigens." CRC Press, Boca Raton, Fla.

Vandenberghe, L. H., Miller, J. R., Gao, G.-P., Calcedo, R., and Wilson, J. M. (2006). The Proposed AAV Serotypes 10 and 11 Serologically Cross-React with AAV8 and AAV4. *Molecular Therapy* **13,** S47-S48.

Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S., and Carragher, B. (2009). DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J Struct Biol* **166**(2)**,** 205-13.

Walsh, C. E., Liu, J. M., Xiao, X., Young, N. S., Nienhuis, A. W., and Samulski, R. J. (1992). Regulated high level expression of a human gamma-globin gene introduced into erythroid cells by an adeno-associated virus vector. *Proc Natl Acad Sci U S A* **89**(15)**,** 7257-61.

Walters, R. W., Yi, S. M., Keshavjee, S., Brown, K. E., Welsh, M. J., Chiorini, J. A., and Zabner, J. (2001). Binding of adeno-associated virus type 5 to 2,3-linked sialic acid is required for gene transfer. *J Biol Chem* **276**(23)**,** 20610-6.

Wang, L., Calcedo, R., Bell, P., Lin, J., Grant, R. L., Siegel, D. L., and Wilson, J. M. (2011). Impact of pre-existing immunity on gene transfer to nonhuman primate liver with adeno-associated virus 8 vectors. *Hum Gene Ther* **22**(11)**,** 1389-401.

Wang, X. S., and Srivastava, A. (1997). A novel terminal resolution like site in the adeno-associated virus type 2 genome. *J Virol* **71**(2)**,** 1140-1146.

Ward, P. (1998). Role of the adenovirus DNA-binding protein in in vitro adeno-associated virus DNA replication. *J Virol* **72**(1)**,** 420-427.

Ward, P., Urcelay, E., Kotin, R., Safer, B., and Berns, K. I. (1994). Adeno-associated virus DNA replication in vitro: activation by a maltose binding protein/Rep 68 fusion protein. *J Virol* **68**(9)**,** 6029-37.

Weindler, F. W., and Heilbronn, R. (1991). A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. *J Virol* **65**(5)**,** 2476-83.

Weitzman, M. D., and Linden, R. M. (2011). Adeno-associated virus biology. *Methods Mol Biol* **807,** 1-23.

Wikoff, W. R., Wang, G., Parrish, C. R., Cheng, R. H., Strassheim, M. L., Baker, T. S., and Rossmann, M. G. (1994). The Structure of a Neutralized Virus: Canine parvovirus Complexed with Neutralizing Antibody Fragment. *Structure* **2**(7)**,** 595-607.

Wistuba, A., Kern, A., Weger, S., Grimm, D., and Kleinschmidt, J. A. (1997). Subcellular compartmentalization of adeno-associated virus type 2 assembly. *Journal of Virology* **71**(2)**,** 1341-1352.

Wistuba, A., Weger, S., Kern, A., and Kleinschmidt, J. A. (1995). Intermediate of Adeno-Associated Virus Type 2 Assembly: Identification of Soluble Complexes Containing Rep and Cap Proteins. *Journal of Virology* **69,** 5311-5319.

Wobus, C. E., Hugle-Dorr, B., Girod, A., Petersen, G., Hallek, M., and Kleinschmidt, J. A. (2000). 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* **74**(19)**,** 9281-93.

Wonderling, R. S., Kyostio, S. R., and Owens, R. A. (1995). A maltose-binding protein/adeno-associated virus Rep68 fusion protein has DNA-RNA helicase and ATPase activities. *J Virol* **69**(6)**,** 3542-3548.

Wu, P., Xiao, W., Conlon, T., Hughes, J., Agbandje-McKenna, M., Ferkol, T., Flotte, T., and Muzyczka, N. (2000). Mutational analysis of the adeno-associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism. *J Virol* **74**(18)**,** 8635-47.

Wu, Z., Asokan, A., and Samulski, R. J. (2006). Adeno-associated virus serotypes: vector toolkit for human gene therapy. *Mol Ther* **14**(3)**,** 316-27.

Wu, Z., Miller, E., Agbandje-McKenna, M., and Samulski, R. J. (2006). Alpha2,3 and alpha2,6 N-linked sialic acids facilitate efficient binding and transduction by adeno-associated virus types 1 and 6. *J Virol* **80**(18)**,** 9093-103.

Xiao, C., and Rossmann, M. G. (2006). Interpretation of electron density with stereographic roadmap projections. *J Struct Biol*.

Xiao, W., Chirmule, N., Berta, S. C., McCullough, B., Gao, G., and Wilson, J. M. (1999). Gene therapy vectors based on adeno-associated virus type 1. *J Virol* **73**(5)**,** 3994-4003.

Xiao, W., Warrington, K. H., Jr., Hearing, P., Hughes, J., and Muzyczka, N. (2002). Adenovirus-facilitated nuclear translocation of adeno-associated virus type 2. *J Virol* **76**(22)**,** 11505-17.

Xiao, X., Li , J., McCown, T. J., and Samulski, R. J. (1997). Gene transfer by adeno-associated virus vector into the central nervous system. *Experimental Neurology* **144**(1)**,** 113-124.

Xie, Q., Bu, W., Bhatia, S., Hare, J., Somasundaram, T., Azzi, A., and Chapman, M. S. (2002). The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc Natl Acad Sci U S A* **99**(16)**,** 10405-10.

Xie, Q., Hare, J., Turnigan, J., and Chapman, M. S. (2004). Large-scale production, purification and crystallization of wild-type adeno-associated virus-2. *J Virol Methods* **122**(1)**,** 17-27.

Xie, Q., Lerch, T. F., Meyer, N. L., and Chapman, M. S. (2011). Structure-function analysis of receptor-binding in adeno-associated virus serotype 6 (AAV-6). *Virology* **420**(1)**,** 10-9.

Zadori, Z., Szelei, J., Lacoste, M. C., Li, Y., Gariepy, S., Raymond, P., Allaire, M., Nabi, I. R., and Tijssen, P. (2001). A viral phospholipase A2 is required for parvovirus infectivity. *Dev Cell* **1**(2)**,** 291-302.

Zaiss, A. K., and Muruve, D. A. (2008). Immunity to adeno-associated virus vectors in animals and humans: a continued challenge. *Gene Ther* **15**(11)**,** 808-16.

Zhang, C., Cortez, N. G., and Berns, K. I. (2007). Characterization of a bipartite recombinant adeno-associated viral vector for site-specific integration. *Hum Gene Ther* **18**(9)**,** 787-97.

Zhao, W., Zhong, L., Wu, J., Chen, L., Qing, K., Weigel-Kelley, K. A., Larsen, S. H., Shou, W., Warrington, K. H., Jr., and Srivastava, A. (2006). Role of cellular FKBP52 protein in intracellular trafficking of recombinant adeno-associated virus 2 vectors. *Virology* **353**(2)**,** 283-93.

Zhong, L., Zhou, X., Li, Y., Qing, K., Xiao, X., Samulski, R. J., and Srivastava, A. (2008). Single-polarity recombinant adeno-associated virus 2 vector-mediated transgene expression in vitro and in vivo: mechanism of transduction. *Mol Ther* **16**(2)**,** 290-5.

Zuffi, E., Manaresi, E., Gallinella, G., Gentilomi, G. A., Venturoli, S., Zerbini, M., and Musiani, M. (2001). Identification of an immunodominant peptide in the parvovirus B19 VP1 unique region able to elicit a long-lasting immune response in humans. *Viral Immunol* **14**(2)**,** 151-8.