Identification and Characterization of Nuclear and Nucleolar Localization Signals in the Adeno-Associated Virus Serotype 2 Assembly-Activating Protein

Lauriel F. Earley,a,b Yasuhiro Kawano,a,c Kei Adachi,b Xiao-Xin Sun,b Mu-Shui Dai,b Hiroyuki Nakaib

Departments of Molecular Microbiology and Immunologya and Molecular and Medical Genetics,b Oregon Health and Science University School of Medicine, Portland, Oregon, USA; Takara Bio Inc., Otsu, Shiga, Japanb

ABSTRACT
Assembly-activating protein (AAP) of adeno-associated virus serotype 2 (AAV2) is a nucleolar-localizing protein that plays a critical role in transporting the viral capsid VP3 protein to the nucleolus for assembly. Here, we identify and characterize AAV2 AAP (AAP2) nuclear (NLS) and nucleolar (NoLS) localization signals near the carboxy-terminal region of AAP2 (amino acid positions 144 to 184) (AAP2144–184). This region contains five basic-amino-acid-rich (BR) clusters, KSKRSRR (AAP2BR1), RRR (AAP2BR2), RFR (AAP2BR3), RSTSSR (AAP2BR4), and RRIK (AAP2BR5), from the amino terminus to the carboxy terminus. We created 30 AAP2BR mutants by arginine/lysine-to-alanine mutagenesis or deletion of AAP2BRs and 8 and 1 green fluorescent (GFP)-AAP2BR and J VI (AAP2BR2), RFR (AAP2BR3), RSTSSR (AAP2BR4), and RRIK (AAP2BR5), from the amino terminus to the carboxy terminus. The results showed that AAP2144–184 has redundant multipartite NLSs and that any combinations of 4 AAP2BRs, but not 3 or less, can constitute a functional NLS-NoLS; AAP2BR1 and AAP2BR2 play the most influential role for nuclear localization, but either one of the two AAP2BRs is dispensable if all 4 of the other AAP2BRs are present, resulting in 3 different, overlapping NLS motifs; and the NoLS is shared redundantly among the five AAP2BRs and functions in a context-dependent manner. AAP2BR mutations not only resulted in aberrant intracellular localization, but also attenuated AAP2 protein expression to various degrees, and both of these abnormalities have a significant negative impact on capsid production. Thus, this study reveals the organization of the intermingling NLSs and NoLSs in AAP2 and provides insights into their functional roles in capsid assembly.

IMPORTANCE
Adeno-associated virus (AAV) has become a popular and successful vector for in vivo gene therapy; however, its biology has yet to be fully understood. In this regard, the recent discovery of the assembly-activating protein (AAP), a nonstructural, nucleolar-localizing AAP protein essential for viral capsid assembly, has provided us a new opportunity to better understand the fundamental processes required for virion formation. Here, we identify clusters of basic amino acids in the carboxy terminus of AAP from AAV serotype 2 (AAV2) that act as nuclear and nucleolar localization signals. We also demonstrate their importance in maintaining AAP expression levels and efficient production of viral capsids. Insights into the functions of AAP can elucidate the requirements and process for AAV capsid assembly, which may lead to improved vector production for use in gene therapy. This study also contributes to the growing body of work on nuclear and nucleolar localization signals.

aden-associated virus (AAV) is a small, single-stranded DNA virus from the parvovirus family that has become a successful vector for gene delivery. The recent achievements in the field of AAV vector research have called attention to the incompletely understood life cycle of the virus. The AAV genome comprises two genes, rep and cap, which encode the nonstructural Rep proteins and the structural VP proteins, respectively. The AAV virion is composed of 60 subunits comprising the three VP proteins, VP1, VP2, and VP3, encoded by a single open reading frame (ORF) in the cap gene. Recently, a nonstructural viral protein encoded by an alternative ORF within the cap gene was identified and termed assembly-activating protein (AAP) for its indispensable role in capsid formation (1–3). The AAP (AAP2) from AAV serotype 2 (AAV2) is a nucleolar-localizing protein that binds to VP proteins through interacting domains in the amino (N) terminus of AAP2 (1), transports the VP proteins to the nucleolus, and promotes capsid assembly (3). Therefore, AAP2 is expected to have both a nuclear localization signal (NLS) and a nucleolar localization signal (NoLS) within its protein sequence. However, such organelle-targeting sequences in AAP2 remain to be identified and characterized.

The most common mechanism for targeting a protein to the nucleus is by an NLS that is recognized by one of the nuclear import proteins, termed importins, which are part of the large family of transport proteins known as karyopherins (4). Classical NLSs can be either monopartite, such as the PKKKRKV sequence
in simian virus 40 (SV40) large T antigen (5), or bipartite, such as the KRPAATKQAGQAKKKK sequence in nucleophosmin (6, 7). These classical NLSs are bound by the adaptor protein importin-α, which is then bound by importin-β, forming a heterotrimeric complex consisting of the two importin proteins and the cargo protein. Importin-β mediates nuclear entry of the heterotrimer through the nuclear pores by its increasing affinity for nucleoporins along the inside of the nuclear pore complex (8). If the cargo protein also contains an NoLS, it can then be targeted to the nucleolus through charge-based interactions (9) or interactions with nucleolar proteins (10, 11), although the specific requirements defining nucleolar localization are not as well understood as those for nuclear import.

As AAP2 is able to localize to the nucleolus (3), we hypothesized that it would contain both an NLS and an NoLS responsible for this intracellular localization and that these signals would be critical to its function in capsid assembly. Because a protein region rich in basic amino acid residues is a hallmark of NLSs and NoLSs, we tested our hypothesis on the carboxy (C)-terminal region of AAP2, amino acid positions 144 to 184 (AAP2144–184), where there are five basic-amino-acid-rich (BR) clusters. By fusing green fluorescent protein (GFP) or the β-galactosidase protein with an AAP2 protein segment of interest and by creating a series of arginine/lysine-to-alanine mutations or deletions in AAP2144–184, we were able to identify NLSs and NoLSs and elucidate their redundant and overlapping nature. Mutations in this NLS- or NoLS-containing region resulted in not only aberrant intracellular localization, but also substantial reduction in AAP2 expression and capsid production, showing the multifaceted functional importance of the NLS-NoLS in AAP2.

MATERIALS AND METHODS
Plasmid construction. pCMV-FLAG-cmAAP2 is a plasmid expressing a codon-modified (cm) version of the wild-type AAP2 with an N-terminal FLAG tag under the control of the human cytomegalovirus (CMV) immediate-early gene enhancer/promoter (12). This CMV-tagged AAP2 is translated from the ATG start codon. The codon-modified AAP2 ORF was utilized to maximize expression in human cells and to prevent recombination between the AA V2-RepVP3 viral genome (12) and AAP2 plasmid DNA through the homologous sequence during AAV production in human embryonic kidney (HEK 293) cells. pCMV-FLAG-cmAAP2 mutant plasmids were created by site-directed mutagenesis. pCMV-GFP is an enhanced GFP (eGFP)-expressing plasmid under the control of the same CMV enhancer/promoter as that in pCMV-FLAG-cmAAP2 and was used to construct plasmids expressing GFP fused with the AAP2144–184 peptide at its C terminus. pAA V2-RepVP3 is a plasmid that expresses all the Rep and VP3 proteins from AAP2 but does not express VP1, VP2, or AAP2 (12). An adenovirus helper plasmid, pHelper, was purchased from Agilent. pCMV-AAV2VP3 is a plasmid expressing the AAV2 VP3 protein under the same CMV enhancer/promoter (12). pAA V-CMV- lacZ is a plasmid containing an alcohol dehydrogenase (Adh)-lacZ fusion transgene under the control of the CMV enhancer/promoter (13). This plasmid expresses cytosolic β-galactosidase in cells transfected with the plasmid. To construct the pAAV-CMV-lacZ-AAP2144–184 plasmid expressing β-galactosidase fused with the AAP2144–184 peptide, the peptide-coding nucleotide sequence was introduced at the N terminus between the start and second codons of the Adh-lacZ gene ORF in the plasmid pAAV-CMV-lacZ. pCMV-FLAG-cmAAP2-GFP is a plasmid expressing the wild-type full-length AAP2 fused with a FLAG tag and GFP at the N terminus and C terminus, respectively. This protein configuration allows simultaneous detection of the wild-type AAP2 in cells using three different approaches: anti-FLAG antibody immunostaining, direct detection of GFP fluorescence, and anti-GFP antibody immunostaining.

Cells. HEK 293 cells (AAV293) were purchased from Stratagene. The HeLa human cervical cancer cell line was obtained from the American Type Culture Collection (ATCC). HEK 293 cells and HeLa cells were grown in Dulbecco’s modified Eagle’s medium (DMEM) (Lonza, Basel, Switzerland) supplemented with 10% fetal bovine serum (FBS), 1-glutamine, and penicillin-streptomycin.

Immunofluorescence microscopy and data analysis. HeLa cells were seeded on coverslips in 12-well plates and transfected with plasmid DNA using polyethyleneimine (PEI). Forty-eight hours after transfection, the cells were fixed with 4% paraformaldehyde at room temperature, permeabilized with 0.2% Tween 20, blocked with 8% bovine serum albumin (BSA), and stained with mouse monoclonal anti-FLAG M2 antibody (F1804; Sigma-Aldrich, St. Louis, MO) and rabbit polyclonal anti-nucleostemin antibody (AB5689 [Millipore, Billerica, MA] or sc-67012 [Santa Cruz Biotechnology, Dallas, TX]), followed by DAPI (4′,6-diamidino-2-phenylindole), Alexa Fluor 488-AffiniPure goat anti-mouse IgG antibody (115-545-166; Jackson ImmunoResearch, West Grove, PA), and Cy3-AffiniPure goat anti-rabbit IgG antibody (111-165-144; Jackson ImmunoResearch). For imaging of AAP2 and VP proteins together, the cells were stained with rat monoclonal anti-DYKDDDDK (FLAG) antibody (NBAP-06712; Novus Biological, Littleton, CO), mouse monoclonal anti-AAV VP1/VP2/VP3 antibody (B1) (03-61058; American Research Products, Inc., Waltham, MA), and rabbit polyclonal anti-nucleostemin antibody (sc-67012; Santa Cruz Biotechnology), followed by DAPI, Alexa Fluor 488-AffiniPure goat anti-mouse IgG antibody (115-545-166; Jackson ImmunoResearch), Cy3-AffiniPure donkey anti-rat IgG antibody (712-165-153; Jackson ImmunoResearch), and Alexa Fluor 647-AffiniPure goat anti-rabbit IgG antibody (111-605-144; Jackson ImmunoResearch). For a subcellular-localization analysis of β-galactosidase, the cells were treated in the same manner as for the double immunostaining of AAP2 and nucleostemin described above, except that mouse monoclonal anti-β-galactosidase antibody (B0271; Sigma-Aldrich) was used in place of mouse monoclonal anti-FLAG M2 antibody. For antigen retrieval by protease treatment, transfected HeLa cells were fixed with 4% paraformaldehyde at 48 h posttransfection, permeabilized with 0.2% Tween 20, incubated with trypsin (2.5 μg/ml) at 37°C for 10 min, treated with 1 mM phenylmethylsulfonyl fluoride (PMSF) at room temperature for 1 min, and immunostained using mouse monoclonal anti-eGFP antibody (F56-6A1.2.3, Thermo Scientific, Waltham, MA) and rat monoclonal anti-DYKDDDDK (FLAG) antibody (NBAP-06712; Novus Biological, Littleton, CO), followed by Alexa Fluor 647-AffiniPure donkey anti-mouse IgG antibody (A-31571; Invitrogen, Grand Island, NY) and Cy3-AffiniPure donkey anti-rat IgG antibody (712-165-153; Jackson ImmunoResearch). The cells were imaged on a Zeiss LSM 710 laser scanning confocal microscope. In this study, we defined NLS (+) proteins as those that showed exclusive nuclear accumulation, NoLS (+) proteins as those that showed obvious nucleolar enrichment, and nucleolar exclusion (Ex) proteins as those that showed clear exclusion from the nucleolus.

The functional role of each AAP2BR in nuclear and nucleolar localization was evaluated in the following manner. The following mutants harboring a pair of AAP2BR mutations showed impaired nuclear trafficking: AAP2mt12, -mt17, -mt21, -mt22, and -mt27. In these 5 mutants, when correction of one AAP2BR mutation back to the wild-type amino acid sequence resulted in an NoLS (AAP2BR) that was corrected has an NLS role. Likewise, we focused on the nucleolar-excluded mutants AAP2mt17, -mt22, and -mt27 harboring a pair of AAP2BR mutations to assess the NoLS role of each AAP2BR. In these 3 mutants, when correction of one AAP2BR mutation back to the wild-type amino acid sequence resulted in an NLS (+) phenotype, we interpreted it as showing that the AAP2BR that was corrected has an NLS role. The NLS roles of a given pair of AAP2BRs were also assessed statistically as described below.

X-Gal staining. HEK 293 cells were seeded on a 6-well plate and transfected with 2 μg of either pAAV-CMV-lacZ or pAAV-CMV-lacZ-AAP2144–184 plasmid DNA. Twenty-four hours after transfection, the...
cells were fixed with 2% formaldehyde-0.2% glutaraldehyde in phosphate-buffered saline (PBS) for 5 min, washed with PBS, and stained with 5 mM FeCl₃(CN)₂–5 mM FeCl₂(CN)₆–2 mM MgCl₂–5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (1 mg/ml) in PBS at 37°C for 1 h.

AAV production and an ELISA specific for intact AAP2 particles. AAV2 VP3-only viral particles were produced by transcomplementation, where VP3 and AAP2 (the wild type or mutant) were expressed in HEK 293 cells from two separate plasmids (12). In brief, HEK 293 cells were transfected with pAAV2-RepVP3, pCMV-FLAG-cmAAP2 (the wild type or mutant), and pHelper using PEI in 6-cm dishes. Forty-eight hours after transfection, the cells were washed in PBS and resuspended in 100 µl of HEPES-buffered saline (pH 7.4). The cells were lysed by three freeze-thaw cycles using a dry-ice–ethanol bath, and the supernatants were collected. The resulting cell lysates containing viral particles were then subjected to an A20 antibody-based intact AAV2 capsid-specific enzyme-linked immunosorbent assay (ELISA) using the AAV2 Titration ELISA Kit (Progen, Heidelberg, Germany) according to the manufacturer’s instructions.

Western blot analysis. HEK 293 cells were transfected with either the wild-type or mutant pCMV-FLAG-cmAAP2 plasmid DNA. Forty-eight hours posttransfection, the HEK 293 cells were lysed in radioimmunoprecipitation assay (RIPA) buffer containing protease inhibitors (Complete Mini; Roche, Indianapolis, IN). Protein concentrations in the cell lysates were determined with a DC Protein Assay Kit (Bio-Rad, Hercules, CA). The same amount of total cell lysates (60 or 80 µg per lane) was separated on an 8% SDS-PAGE gel, transferred onto a polyvinylidene difluoride (PVDF) membrane, and reacted with mouse monoclonal anti-FLAG M2 antibody and monoclonal mouse anti-α-tubulin antibody (sc-32293; Santa Cruz Biotechnology), followed by goat polyclonal anti-mouse IgG antibody (sc-2055; Santa Cruz Biotechnology) conjugated to horseradish peroxidase. The signals on the blots were visualized and quantified using the FluorChem M system (ProteinSimple, Santa Clara, CA). The data were collected from a biologically duplicated set of experiments.

Statistical analyses. Differences in AAV particle production yields were statistically assessed by the two-tailed Welch t test. An unconditional exact test was used to statistically evaluate the association between the presence or absence of a given combination of 2 intact AAP2BRs and the presence or absence of an NLS using a 2-by-2 contingency table. For the purpose of this statistical analysis, the NLS (+) and NLS (−) mutants among AAP2mt10 to -mt27 were defined as those that showed exclusively nuclear accumulation and those that did not belong to the NLS (+) category, respectively. There were 9 NLS (+) and 9 NLS (−) AAP2BR mutants. The null hypothesis is that there is no association between AAP2BRs and NLSs. Since we analyzed only 18 combinations out of all 25 possible combinations of AAP2BR mutations that leave at least 2 intact AAP2BRs (10 combinations of 2 intact BRs with 3 mutated BRs, 10 combinations of 3 intact BRs with 2 mutated BRs, and 5 combinations of 4 intact BRs with 1 mutated BR), a range of P values are given in which all the possible outcomes from the 7 AAP2BR mutants that were not assessed are taken into account. The P values that would be obtained if all 25 AAP2BR mutants were analyzed are expressed as P_{obs}. The P values obtained from the 18 AAP2BR mutants are expressed as P_{18}.

RESULTS

The C terminus of AAP2 contains both an NLS and an NoLS. AAP2 has an amino acid stretch rich in basic amino acid residues, AAP2144−184, near the C terminus (Fig. 1). This region harbors a KRKR sequence that matches the classical monopartite NLS motif, K(ER)/R/X(K/R) (14), and a putative NoLS, RRIK (15, 16), and shares features characteristic of NoLSs, including a high proportion of basic amino acids and proximity to the C terminus (17). It is also possible that these two basic amino acid clusters might serve as a bipartite NLS with a long linker sequence (14, 18, 19). Therefore, we hypothesized that AAP2144−184 contains both an NLS and an NoLS. To test this hypothesis, we transfected HeLa cells with the pCMV-GFP-AAP2144−184 plasmid expressing the GFP fused with the AAP2144−184 peptide at its C terminus or with pCMV-GFP, the control parental GFP plasmid devoid of the AAP2144−184 peptide. Forty-eight hours posttransfection, the control GFP was observed diffusely throughout the cytoplasm and nucleoplasm, while the GFP-AAP2144−184 fusion protein was localized predominantly in the nucleus (Fig. 2A and B). Since GFP can diffuse into the nucleus without an NLS due to its small size (27 kDa), we also investigated whether the AAP2144−184 peptide could target a large cytoplasmic protein, β-galactosidase (~120 kDa), to the nucleus. Both immunofluorescence microscopic and cytochemical analyses demonstrated strong nuclear enrichment of the otherwise predominantly cytosolic β-galactosidase when the enzyme was fused with the AAP2144−184 peptide (Fig. 2D). Addition of the canonical NLS derived from SV40 large T antigen (PKKKRKV) at the N terminus of β-galactosidase was not sufficient for nuclear targeting, showing a nuclear-exclusion pattern, although the protein accumulated in the nucleus (data not shown). These observations strongly supported our hypothesis that the AAP2144−184 region contains both an NLS and an NoLS.

FLAG-tagged wild-type AAP2 and nucleostemin colocalize in the nucleus and nuclear bodies. It has been established that
Au1-tagged AAP2 and fibrillarin, an endogenous nucleolar protein, colocalize in the nucleolus in HeLa cells (3). We first investigated whether the N-terminally FLAG-tagged AAP2 we used for this study also localizes in the nucleus as expected. To this end, HeLa cells were transfected with pCMV-FLAG-AAP2144–184 and pCMV-FLAG-cmAAP2 plasmid (Fig. 1). Each mutant AAP2 was expressed in HeLa cells by plasmid DNA transfection and analyzed for its intracellular localization by immunofluorescence microscopy (Fig. 3). We first created AAP2 mutants 1 to 12 (Fig. 1, mt1 to mt12) to investigate the roles of AAP2BR1 and AAP2BR5, which carry the classical monopartite NLS motif, KRKR, and the putative NoLS motif, RRK, respectively (basic amino acid residues in the motifs are underlined). Removal of positive charges from AAP2BR1 (AAP2mt1 to -mt10) or from AAP2BR5 (AAP2mt11) did not abolish the ability of mutant AAP2s to translocate to the nucleus with strong nucleolar enrichment and robust accumulation in the nucleostemin-positive nuclear bodies (Fig. 3B and C). Removal of positive charges from AAP2BR1 and AAP2BR5 in conjunction (AAP2mt12), however, resulted in substantial, if not complete, nuclear exclusion and nucleolar exclusion (Fig. 3D). This first set of experiments indicated that there are at least two NLSs that involve either AAP2BR1 or AAP2BR5 within AAP2144–184. We then created an additional set of 18 AAP2 mutants, 13 to 30 (Fig. 1, mt13 to mt30), to further investigate the role of each AAP2BR in nuclear and nucleolar localization. In these mutants, the positive charge in one or more AAP2BRs was diminished by alanine substitutions or deletions in various combinations. The 7 AAP2 mutants that had only 1 mutant AAP2BR and 4 intact AAP2BRs (mt10, mt11, mt13, mt14, mt15, mt28, and mt29) were all found exclusively in the nucleus with obviously enhanced association with the nucleolus (Fig. 3B, C, E to G, T, and U). In keeping with this observation, GFP-AAP2144–184 fusion mutants that had four intact AAP2BRs were observed exclusively in the nucleus, with strong enrichment in the nucleolus (Fig. 5A to C). All 13 mutants that had 3 or fewer intact AAP2BRs exhibited attenuated or complete loss of nucleolar association, and among them, 9 (mt16, mt17, mt20, mt21, mt22, mt24, mt25, mt27, and mt30) showed various degrees of impaired nuclear transport and cytoplasmic retention (Fig. 3D and H to V). AAP2 is a small protein that can passively cross the nuclear envelope through the nuclear pore complexes. The cytoplasmic signals that are stronger than the nuclear signals in AAP2mt12, -mt20, and -mt24, therefore, might indicate the presence of a nuclear export signal (NES) in AAP2. However, we have not observed any nuclear-cytoplasmic shuttling of the AAP2 protein during the course of AAV2 particle production in HEK 293 cells by plasmid transfection, where AAP2 consistently remains located in the nucleolus or nucleostemin-positive nuclear bodies (L. F. Earley and H. Nakai, unpublished observation). In addition, three different NES prediction programs (21–23) failed to identify any potential CRM1-binding NES in the AAP2 amino acid sequence. Thus, the above-mentioned observations are most likely due primarily to various degrees of impairment of nuclear import caused by different mutations, although the involvement of an unidentified NES cannot be totally ruled out. Taken together, these observations indicate that the NLSs and NoLSs in AAP2 are redundant and that various combinations of 4 AAP2BRs out of 5 AAP2BRs within the AAP2144–184 region are able to constitute a functional multipartite NLS-NoLS.
Copresence of AAP2BR1 and AAP2BR2 is an important but not absolute requirement for the AAP2 NLS. The immunofluorescence microscopy analysis described above revealed that all 10 AAP2BR mutants that exhibited impaired nuclear translocation and various degrees of cytoplasmic retention had either or both AAP2BR1 and AAP2BR2 mutations, and all 10 combinatorial mutations involving AAP2BR1 or AAP2BR2 impaired nuclear transport of AAP2 to various degrees. A statistical comparison of various combinations of two intact AAP2BRs and the presence of a fully functional NLS in the AAP2 mutants revealed that the combination of AAP2BR1 and AAP2BR2 is strongly associated with nuclear localization (Table 1). The GFP fusion experiment also demonstrated that AAP2BR1 by itself is not sufficient for active nuclear transport and that copresence of AAP2BR1 and AAP2BR2 is required for nuclear accumulation (Fig. 5D to H). When we assessed the NLS role of each AAP2BR based on its ability to restore a fully functional NLS to AAP2 mutants (see Materials and Methods), we found that all 5 AAP2BRs can contribute to the facilitation of nuclear localization of AAP2. These observations indicate that, although all 5 AAP2BRs have an NLS role, the presence of AAP2BR1 and AAP2BR2 in conjunction plays the most influential role in nuclear targeting. Due to the redundant nature of the AAP2 NLS, either one of these two AAP2BRs is dispensable if all 4 of the other AAP2BRs are present, as evidenced by the nuclear localization of AAP2mt10 and -mt13. All the AAP2BRs can contribute to nucleolar localization and function in a context-dependent manner. The NoLS role of each AAP2BR was investigated based on the functional restoration of nucleolar-excluded AAP2 mutants, as detailed in Materials and Methods. This analysis revealed an obvious NoLS role in AAP2BR1, -2, -4, and -5. For example, the NoLS role of AAP2BR2 could be demonstrated by the observation that, when the mutated
AAP2BR2 was corrected back to the wild-type sequence in the nucleolar-excluded AAP2mt17 (Fig. 3I), nucleolar retention was restored, as evidenced with AAP2mt11 (Fig. 3C). A comparison between AAP2mt19, showing modest nucleolar accumulation, and AAP2mt26, showing nucleolar exclusion, which differ only by the presence or absence of AAP2BR3, revealed that AAP2BR3 also functions as a weak NoLS (Fig. 3K and R). These results indicate that effective nucleolar accumulation is a consequence of the co-operation of the redundant NoLSs residing in the five AAP2BRs in a given combination of AAP2BRs. Interestingly, AAP2mt26 (AAP2BR1+/BR2+/BR3+/BR4-/BR5−), where basic amino acids in the BRs are retained [plus] or mutated to alanine [minus]) exhibited nearly complete nuclear localization with nucleolar exclusion (Fig. 3R), while GFP-AAP2144–184mt26, carrying AAP2144–184 with the same AAP2BR mutations, and GFP-AAP2144–156, in which GFP was fused with a 13-mer peptide containing only AAP2BR1 and AAP2BR2, showed nuclear and nucleolar accumulation (Fig. 5F and H). The same discrepancy in the pattern of nucleolar accumulation was observed in AAP2mt22 (Fig. 3N) and GFP-AAP2144–184mt22 (Fig. 5E). It has previously been shown that adding a nonapeptide containing 9 basic amino acids to the C terminus of GFP allows nucleolar accumulation of the protein, not through a specific NoLS motif, but presumably through nonspecific, electrostatic interactions with negatively charged nucleolar components (9). There are 8 basic amino acids within the AAP2BR1-AAP2BR2 segment, and there are 10 basic residues in AAP2mt22 (Fig. 1). Therefore, the extra positive electric charges added to GFP likely explain why the GFP-AAP2144–156, GFP-AAP2144–184mt22, and GFP-AAP2144–184mt26 fusion proteins were still capable of nucleolar retention while AAP2mt26 was not. Taken together, although there is no inconsistency in the NLS roles of AAP2BR1 and AAP2BR2 between the AAP2 mutants and the GFP fusion proteins, their NoLS roles are context dependent.

**Capsid assembly-promoting functions of AAP2BR mutants.**
A series of studies have demonstrated that nucleolar localization of viral components is important for the AAV life cycle (24–27). Indispensable roles of the nucleolus in the virus life cycle have also been demonstrated in other viruses (28–30). Having created a panel of AAP2 mutants whose intracellular localizations are characterized, we addressed how aberrant localization of AAP2 might affect its assembly-promoting function. To this end, we per-

---

**FIG 4** Immunofluorescence microscopic analysis of a FLAG-tagged AAP2 fused with GFP in HeLa cells with or without proteinase treatment. (A to C) HeLa cells were transiently transfected with pCMV-FLAG-cmAAP2-GFP expressing the wild-type full-length AAP2 fused with a FLAG tag and GFP. Intracellular localization of the fusion protein was analyzed with or without trypsin treatment by anti-FLAG antibody immunostaining (A), direct detection of GFP fluorescence (B), and anti-GFP antibody immunostaining (C) under a fluorescence microscope. (D) Merged images. Scale bar, 20 μm.

**FIG 5** Microscopic assessment of the abilities of the AAP2BRs to target GFP to the nucleus and the nucleolus. The method used was the same as that for Fig. 2. (A to F) GFP-AAP2144–184mt13 to -mt26 are GFPs fused with the 41-amino-acid-long AAP144–184 region containing their corresponding AAP2BR mutations. (G and H) GFP-AAP2144–150 and GFP-AAP2144–156 are GFPs fused with AAP2BR1 only (7 amino acids) and AAP2BR1-AAP2BR2 only (13 amino acids), respectively. Scale bar, 20 μm.
TABLE 1

<table>
<thead>
<tr>
<th>AAP2BR1</th>
<th>AAP2BR2</th>
<th>AAP2BR3</th>
<th>AAP2BR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001-0.017</td>
<td>0.0001-0.00001</td>
<td>0.01-0.1</td>
<td>0.1-1</td>
</tr>
<tr>
<td>0.01-0.1</td>
<td>0.1-1</td>
<td>1-10</td>
<td>10-100</td>
</tr>
<tr>
<td>0.1-1</td>
<td>1-10</td>
<td>10-100</td>
<td>100-1000</td>
</tr>
<tr>
<td>1-10</td>
<td>10-100</td>
<td>100-1000</td>
<td>1000-10000</td>
</tr>
<tr>
<td>10-100</td>
<td>100-1000</td>
<td>1000-10000</td>
<td>10000-100000</td>
</tr>
</tbody>
</table>

Eighteen AAP2 mutants (AAP2mt10 to -mt27) were categorized into two groups according to the presence or absence of an NLS as defined in the legend to Fig. 1. There were 9 NLS (+) and 9 NLS (−) AAP2 mutants. An unconditional exact test was performed to statistically evaluate the association between the presence or absence of a given combination of 2 AAP2BRs and the presence of an NLS. The null hypothesis is that there is no association between the presence or absence of a given combination of 2 AAP2BRs and the presence of an NLS. The null hypothesis of no association is rejected at a significance level of 0.05.

The statistical analysis for the combination does not give a conclusive result due to the wide range of P values across the statistically significant cutoff value of 0.05.

The data presented above provided insights into the AAP2BRs' role in this respect well explains why some of the mutants failed to produce capsids effectively.

**AAP2BRs maintain AAP2 expression levels for effective capsid assembly.** During the microscopic analysis, we had noticed that the numbers of transfected HeLa cells and the signal intensities in many AAP2 mutants were lower than those from the wild type to various degrees. Since AAP2BR mutations may have affected not only intracellular localization, but also protein expression levels, we performed a quantitative Western blot assay on lysates from HEK 293 cells transfected with either the wild-type or mutant pCMV-FLAG-cmAAP2 plasmid. The molecular mass of the FLAG-tagged AAP2 protein was determined to be 27 kDa by SDS-PAGE (26 kDa for AAP2), which is slightly higher than a calculated molecular mass of 24 kDa. Mutation or deletion of the AAP2BR regions decreased AAP2 protein expression levels to various degrees between 0.02 and 0.64 compared to 1.0 for the wild type (Fig. 6A). The expression levels of all the mutants showing aberrant localization were reduced substantially, with the relative levels ranging from 0.02 to 0.14. The reduction of protein levels was particularly pronounced in the nucleolar-excluded mutants. The AAP2BR1 or AAP2BR2 mutants showing localization similar to that of the wild type had greater quantities of protein on average than the mutants showing aberrant localization; however, they also showed attenuated AAP2 expression ranging from 0.03 to 0.64. Importantly, the results showed a very nice correlation between intracellular AAP2 levels, nucleolar accumulation of AAP2, and AAV capsid production. When AAP2 mutants retained the ability to localize normally (these mutants are indicated by the horizontal black line in Fig. 6A), AAV capsid yields were positively correlated with the AAP2 mutant protein levels, with a Pearson correlation coefficient of 0.94 (Fig. 6B). When AAP2 mutants were excluded from the nucleolus, AAV capsid yields were disproportionately lower than the AAP2 mutant protein levels (these mutants are indicated by the horizontal gray line in Fig. 6A). Although the underlying mechanism has yet to be elucidated, these results indicate that AAP2BRs play a pivotal role in maintaining the AAP2 levels for effective capsid assembly. The AAP2BRs’ role in this respect well explains why some of the AAP2BR1 or AAP2BR2 mutants showing wild-type localization failed to produce capsids effectively.

**The AAP2 NoLS is important for nucleolar accumulation of AAV2 VP3.** The data presented above provided insights into the AAP2BRs’ functional roles in intracellular localization of AAP2
AAP2 role as a VP3 transporter. Since the nucleolus has been thought to be important for AAV2 capsid assembly and nucleolar-excluded AAP2BR mutants showed a substantially reduced ability to produce viral capsids, it is possible that AAP2BR mutants would not associate with VP3, leading to impaired nuclear transport and/or failure to accumulate VP3 in the nucleolus even if VP3 is actively localized to the nucleus. To address this, HeLa cells were transfected with pCMV-AAV2VP3 and pCMV-FLAG-cmAAP2 (either the wild type or AAP2mt26), and intracellular localization of VP3 and VP3 was analyzed by immunofluorescence microscopy at 48 h posttransfection. As Sonntag et al. have previously shown, VP3 diffusely localizes to the cytoplasm and nucleoplasm but remains outside the nucleus unless AAP2 is present (3) (Fig. 8A and B). In the presence of AAP2mt26, which had the ability to exclusively translocate to the nucleus but was not able to accumulate in the nucleolus, VP3 was seen predominantly in the nucleoplasm, colocalizing with AAP2mt26 and leaving only a small quantity of VP3 proteins in the cytoplasm. Importantly, VP3 failed to accumulate in the nucleolus (Fig. 8C). We have also found, using a plasmid carrying an AAV2 cap gene that does not express AAP2, that expression of the AAV2 VP1, VP2, and VP3 proteins does not result in nucleolar localization of any of the VP proteins in HeLa cells when AAP2 is not coexpressed (Earley and Nakai, unpublished). Taken together, these observations indicate that the NoLS in AAP2 plays a primary role in targeting capsid VP proteins to the nucleolus.

![FIG 6](image-url) AAV2 VP3 capsid production titers and wild-type and mutant AAP2 protein expression levels in HEK 293 cells. HEK 293 cells seeded on 6-cm dishes were transfected with pAAV2-RepVP3, pCMV-FLAG-cmAAP2 expressing either the full-length wild-type AAP2 or a full-length AAP2 mutant with an AAP2BR mutation(s), and pHelper, using PEI (12). The cells were harvested at 48 h posttransfection and made into a 100-µl crude cell lysate in HEPES-buffered saline after three cycles of freezing and thawing. The AAV2 particle concentration in each crude lysate was quantified by an A20 antibody-based ELISA. (A) Viral titers relative to the titer obtained with the wild-type AAP2 (black bars). A relative titer of 1.0 corresponds to \(7.2 \times 10^{12}\) particles/ml. The thick horizontal black line between the graph and the x axis indicates the wild type and AAP2 mutants that showed exclusive nuclear localization with enhanced nucleolar association. The thick horizontal gray line indicates the mutants showing nucleolar exclusion. The data were collected from biologically triplicate experiments. *, P < 0.05 (two-tailed Welch’s t test) compared to the wild-type values. In a separate experiment, HEK 293 cells seeded on 6-cm dishes were transfected with pCMV-FLAG-cmAAP2 (the wild type or mutant) using PEI. The cells were harvested 48 h posttransfection, and mutant AAP2 protein expression levels relative to the level of the wild type were determined by a quantitative Western blot analysis using biologically duplicate samples (gray bars). The error bars represent standard errors of the mean (SEM) (black bars) or the difference between each value and the mean value (gray bars). (B) Correlations between AAP2 protein levels of the wild type and mutants indicated by the thick horizontal black line in panel A and AAV2 VP3 capsid production titers are shown in a scatter plot. All values are relative to the levels of the wild type.

![FIG 7](image-url) Western blot analysis of the wild-type and mutant AAP2 proteins expressed in HEK 293 cells. Cells seeded on 6-cm dishes were transfected with pCMV-FLAG-cmAAP2 (wild type or mutant) using PEI. The same quantity of crude lysates prepared at 48 h posttransfection were separated on an 8% SDS-PAGE gel and blotted onto a membrane, which was then probed with both anti-FLAG and anti-α-tubulin antibodies. The AAP2 mutant numbers are shown above the lanes.

![FIG 8](image-url) Intracellular localization of AAV2 VP3 in the presence or absence of wild-type AAP2 or in the presence of the nuclear-targeted, nucleolar-excluded AAP2 mutant, AAP2mt26. HeLa cells were transfected with the following plasmids: pCMV-AAV2VP3 only (A), pCMV-AAV2VP3 and pCMV-FLAG-cmAAP2 (wild type) (B), and pCMV-AAV2VP3 and pCMV-FLAG-cmAAP2mt26 (C). The cells were fixed at 48 h posttransfection and immunostained with anti-FLAG, anti-AAV VP1/VP2/VP3, and anti-nucleostemin antibodies. Scale bar, 20 µm.
**DISCUSSION**

The NLS in AAP2 had remained unidentified in a series of studies reported by Kleinschmidt’s group (1-3), which prompted us to seek to determine the amino acid residues responsible for nuclear translocation of AAP2. The consensus sequences of the classical, importin-α-binding NLSs have been well defined. The monopartite NLSs have the consensus sequence K(R)/K)X(K/R), while the bipartite NLSs show a consensus sequence composed of two short basic amino acid clusters separated by a linker sequence, such as (K/R)(K/R)X_{10-12}K/R)13/5, where the C-terminal side cluster contains at least three basic amino acids within 5 consecutive residues (14, 18, 19). Many noncanonical, importin-α-binding NLSs have also been identified, including bipartite NLSs with a long linker of up to 29 residues (14, 18, 19). In light of this knowledge about various types of NLSs, one could assume that AAP2 has one classical monopartite NLS in AAP2BR1 (KSKR5RR) and/or one noncanonical bipartite NLS with a long linker in AAP2BR1-X30-AAP2BR5 (RRIK). Our experimental data, however, do not support this assumption. None of the AAP2BRs were capable of serving as a monopartite NLS. Substantial nuclear enrichment required the presence of AAP2BR1-AAP2BR2 (KSKR5RR-X7-RRR) or at least four AAP2BRs when either AAP2BR1 or AAP2BR2 is absent (i.e., AAP2BP1-AAP2BP3-AAP2BP4-AAP2BP5 and AAP2BP2-AAP2BP3-AAP2BP4-AAP2BP5). Thus, we identified at least three different motifs that can function as an NLS and overlap within the AAP2144-184 region. This redundant nature of the AAP2 NLS explains why a previous investigation concluded that AAP2BR1 was not involved in nuclear localization even though the region contained a classical NLS (1).

In nucleolar-localizing proteins, an NLS and an NoLS can reside in separate locations, but they often coexist in a region highly rich in lysine and arginine residues. The NLS and NoLS of AAP2 that we identified is a joint NLS-NoLS, because all the AAP2BR regions play either, or both, of the NLS and NoLS roles. Many such joint NLS-NoLSs have been reported as NLSs in the literature, mainly due to incomplete investigation for nucleolar accumulation, and therefore, joint NLS-NoLSs have been overlooked in many instances (17). A recent attempt to extract commonalities of 48 human NoLSs, including joint NLS-NoLSs, has revealed that NoLSs are highly basic; are predominantly located near the N or C terminus of proteins, in α-helices or coils and rarely in β-strands; and are solvent accessible (17). In this regard, the AAP2144-184 region containing the joint NLS-NoLS is consistent with these general trends. That is, in addition to the high proportion of basic amino acids, the NLS-NoLS resides near the C terminus; the AAP2144-184 residues are exclusively in α-helices (49%) or coils (51%) according to the Jpred3 prediction (31); and the proportions of buried residues in the AAP2144-184 region predicted by Jpred3 are 37%, 10%, and 2% at prediction thresholds of 25%, 5%, and 0%, respectively, which are lower than in the full-length protein control in the study reported by Scott et al. (17). The AAP2144-184 region also contains the previously identified (K/R) (K/R)(K/R) NoLS motif (15). Through a literature search, we could identify a joint NLS-NoLS that resembles the AAP2 NLS-NoLS. Liu et al. reported that a stretch of 18 amino acid residues, 2746KKMKHHKNKSEAKKRI2761, is part of a joint NLS-NoLS found in 1A6/downregulated in metastasis (1A6/DRIM), a large nucleolar-targeting protein (2,785 amino acids) involved in rRNA processing (32). This amino acid sequence is reminiscent of 25R RRRITLPARERCLLTRSTSSRTSSARRIK, one of the complete joint NLS-NoLSs that we identified in AAP2. Although the observations obtained from the arginine/lysine-to-alanine mutagenesis experiments were not exactly the same in the Liu et al. study and ours, both identified basic amino acid clusters that play a dual NLS-NoLS role. In addition, both studies showed that a decrease in electric charge in the signals resulted in nucleolar exclusion. As demonstrated with AAP2 and many other nucleolar-localizing proteins (17), NLSs and NoLSs are inherently difficult to clearly define, particularly when they reside within the same region with a portion of the signal playing a dual role.

How NoLSs perform their role in nucleolar targeting has not been fully understood. Unlike NLSs, NoLSs do not have well-defined consensus sequences, although they share common characteristics to some degree, as described above. The nucleolus does not have any compartmentalizing membranous structures, and therefore, nucleolar accumulation presumably does not require sophisticated machinery that actively targets the organelle, like the nuclear transport machinery. Rather, nucleolar accumulation of proteins has been thought to be mediated by a retention mechanism through interactions with nucleolar components, such as RNA and other nucleolar proteins (33). RNA-binding motifs have been identified in many nucleolar proteins, including nucleolin, fibrillarin, and viral RNA-binding proteins (34-36). Nucleophosmin, a putative shuttle protein that transports cargos between the cytoplasm and the nucleolus (37-40), has been shown to mediate nucleolar retention of other proteins through the interaction between two highly acidic regions (_120EDAESDEDEEEED_132 and _141DDDDEEDDDDDDEDDDEEAEF_159_ in nucleophosmin and positively charged amino acid regions in the other proteins (40). Thus, nucleolar accumulation is most likely mediated by electrostatic interaction rather than by a defined set of particular amino acid sequences (9). Concordant with this notion, we could only vaguely define the AAP2 NoLS, as opposed to its NLSs, and found that the abundance of positively charged AAP2BR regions, rather than a specific amino acid sequence or a specific combination of AAP2BRs, confers on AAP2 the ability to translocate from the nucleoplasm to the nucleolus. This is in line with the notion that the nucleolar accumulation of AAP2 results from nonspecific interaction with negatively charged nucleolar components and also raises the possibility that the degree of accumulation is determined by the net electric charge of the AAP2144-184 region. The electric charge of AAP2144-184 is 15, and at least 12 has been found to be required for efficient nucleolar accumulation of the AAP2BR mutants (Fig. 1). The net electric charges of the AAP2144-184 corresponding regions of various serotype AAPs (AAP1 to -13) (1) vary from 8 to 15, with AAP5 being the lowest and AAP2 and AAP3B being the highest. Assuming that nucleolar accumulation is determined by the abundance of the net positive charges within this region, AAP5 would likely be devoid of an NoLS and excluded from the nucleolus. If this is true, AAP5 would not be able to transport AAV5 capsid VP proteins to the nucleolus, the organelle that is believed to be important for AAV capsid assembly, at least for AAV2 (24-27). How would AAV5 viral capsid proteins assemble if they are not transported to the nucleolus? Although our study demonstrates that the nucleolar localization of AAP2 strongly correlates with the ability of AAV2 to form capsids, it is intriguing to speculate that the nucleolar accumulation of AAP2 might merely be a coincidental, nonspecific consequence
associated with a function(s) other than nucleolar targeting and that this targeting might not be functionally important. Such nonspecific nucleolar accumulation with no functional role has been demonstrated in the human histone H2B protein (9). Future studies on the subcellular localizations of AAPs derived from various serotypes will address this potential paradox and elucidate the biological significance of the presence of an NoLS in AAPs. Nonetheless, our observations corroborate the driving role of the electric charge in nucleolar targeting and support the idea that NoLSs pay a lower cost for the acquisition of their role by intermingling their signals with NLSs that already have high electric charges. This is particularly advantageous to small viruses that have a limited capacity to store genetic information.

Surprisingly, we could identify at least three complete NLSs that overlap within the same AAP2144–184 region. They are AAP2BR1 plus -BR2, AAP2BR1 plus -BR3 plus -BR4 plus -BR5, and AAP2BR2 plus -BR3 plus -BR4 plus -BR5. Redundant NLSs, as either separate NLSs or overlapping NLSs, have been identified in many proteins, including those derived from viruses. For example, the polyomavirus large T antigen has two separate NLSs, VPARKRVR and PKKARED, each of which mediates nuclear localization in the absence of the other (41,42); the nonstructural 5A (NS5A) protein has a 40-residue-long region where two karyopherin-binding motifs, an importin-α-binding motif and a transportin 3-binding motif, overlap (44).

Utilization of different nuclear transport pathways through the interactions between an NLS and multiple karyopherins have also been demonstrated in human immunodeficiency virus (HIV) Rev protein (45). The straightforward inference about the NLS redundancy is that it ensures and promotes efficient nuclear entry by increasing the avidity for the nuclear transport system and/or by using more than one nuclear entry pathway. However, there are also many instances in which the amino acid residues composing an NLS perform functions distinct from those in nuclear translocation. For example, the unconventional NLS in the nucleoprotein (NP) of the influenza A virus plays a crucial role in nuclear import of the viral genomic RNA (46), and the second NLS in the transcription factor Flt-1 also functions as a DNA-binding domain (47). In fact, DNA-binding domains frequently overlap NLSs (47), and this has led to speculation about the possible co-evolutionary origins of the two motifs (48,49). In the case of AAP2, our present and previous studies provide clues to understanding the necessity of redundancy. In our previous study, we used a directed- evolution approach to experimentally and computationally evolve functionally competent mutant AAPs from a randomly mutagenized library (12). We demonstrated that the functionally optimal AAP2 mutants harbor a strongly basic and hydrophilic heptapeptide, (R/S)7, in the AAP2BR1 region that remarkably resembles the native sequence, KSKRSRR (12). Since all the AAP2 mutants used in the previous study had a completely optimal AAP2 amino acid sequence able to specify nuclear localization, Cell 39:499–509. http://dx.doi.org/10.1016/0092-8674(84)90457-4.


5. Wang Y, Chen B, Li Y, Zhou D, Chen S. 2011. PNRC accumulates in the nucleolus by interaction with B23/nucleophosmin via its nucleolar local-


