Adeno-associated Virus Replication Initiator Assembly

Fred Dyda, Alison B. Hickman, Donald R. Ronning, Robert M. Kotin *Laboratory of Molecular Biology, NIDDK, NIH, Bethesda MD, 20892 U.S.A.*

Adeno-Associated Virus, Replication, Gene Therapy

The Rep protein of the adeno-associated virus (AAV) is essential both for viral replication and locus-specific integration of the viral genome into AAVS1, a silent region of human chromosome 19. Locus-specific integration, a property unique to AAV, is a desired property of vector systems used in gene therapy to establish safe and long term expression of the transgene. To understand the molecular mechanisms of AAV replication and integration, we have solved the crystal structure of the endonuclease domain of Rep, the first three-dimensionally characterized member of the Rolling Circle Replication superfamily We have also determined the structures of this domain of Rep bound to its DNA recognition element, or Rep Binding Site (RBS), in the viral inverted terminal repeat (ITR), and bound to a stem-loop of the ITR that is a key regulatory element in replication initiation. At the RBS, five Rep monomers bind to five direct GCTC tetranucleotide repeats, and each repeat is recognized by two Rep monomers from opposing sides of the DNA. At the same time, each Rep monomer recognizes two adjacent repeats, giving rise to a daisy-chain like assembly of Rep monomers on the DNA. The binding of two adjacent repeats by one molecule explains why Rep can tolerate substitutions (imperfect repeats) if a "perfect" GCTC is available next to it. The recognition of several direct repeats by multiple protomers spiraling around the axis of the DNA is an unusual mode of DNA binding. The same domain recognizes the ITR stem-loop structure using a completely different binding surface that interacts with the tip of the loop and recognizes a TxT motif. One of these bases is completely flipped out of the DNA helix and sits in pre-formed pocket of the protein where it is recognized using all possible hydrogen bond donors and acceptors. Taken together, the two complex structures suggest how the nuclease domain of Rep initiates hexameric helicase assembly on the ITR. To verify this, we have used analytical size-exclusion chromatography to show that on authentic ITR sequences the full length protein forms 6:1 complexes. Once assembled, the active helicase unwinds a region of the ITR so that the nuclease domain can introduce a strand and site specific nick, liberating a 3’ OH group that can serve as the replication origin. Given its small size (21kD), the Rep endonuclease domain is remarkable in that it features three distinct DNA recognition sites, the two DNA binding sites and the nuclease active site. This is the structural manifestation of the economy that Rep has to observe given the very limited coding space available in the viral genome and the many functions that Rep must perform in viral replication.