

Élő Péter

Egy atadenovírus izolátum genomjának
összehasonlító elemzése

Ph.D. értekezés tézisei

Témavezetők: dr. Benkő Mária és dr.Harrach Balázs

MTA Állatorvostudományi Kutatóintézete

Budapest
2002

1. fejezet

Tézisek

1.1. Bevezetés és célkitűzések

Az 1950-es években felfedezett adenovírusok a gerincesek kutatási szempontból egyik legjelentősebb csoportját alkotják. Attól a megfigyeléstől kezdve, hogy emberi adenovírusok újszülöt rágcsálókba oltva azokban tumorokat indukálhatnak, a molekuláris biológiai kutatások homlokterébe kerültek mint a tumorképződés modellezésére alkalmasnak látszó objektumok. A rajtuk kifejlesztett vizsgálati eljárások jelentős része napjainkig használatos. Amlett, hogy az adenovírusok transzkripcióis folyamatainak kutatása vezettek el az RNS-érési folyamatok felfedezéséhez, a DNS-replikáció és a vírusrészecskék összeépüléséi mechanizmusainak kutatása terén is modell szerepet töltek be. Az utóbbi évtizedekben pedig a sejtélettani folyamatokat, és az immunrendszert moduláló hatásaik kutatására helyeződött át a hangsúly. A vizsgálatok egy másik vonulata az adenovírusoknak mint potenciális génterápiás vektoroknak a felhasználására irányult, de készültek kutatásban használatos expressziós vektorok is adenovírus alapra. Az adenovírusok kutatásával foglalkozó irodalom ma már könyvtárnyi méreteket ér el.

Ennek ismeretében meglepő, hogy mindezen ismeret kevés kivételtől eltekintve mindössze néhány adenovírus törzsön végzett kutatás eredménye, sőt messze a legtöbb vizsgálatot két közel rokon emberi vírustörzsön a HAdV-2-n és HAdV-5-ön. végezték. A DNS szekvenálás elterjedéséig az állati adenovírusoknál az esetleíráson túl többnyire megelégedtek a szerológiai jellemzéssel, vagy új gazdafaj esetén elektronmikroszkópos azonosítással. Annak ellenére, hogy az 1970-es évek óta ismertek azok az adenovírus típusok, amelyeknek az adott, két nemzetségből (*Aviadenovirus Mastadenovirus*) álló rendszerbe jellemzőik alapján nem sorolhatók be, kellő mennyiségű molekuláris adat hiányában a rendszer megváltoztatására a közelmúltig nem nyílt lehetőség.

Ahogy a tömeges DNS-szekvenálás a nyolcvanas évek vége felé elterjedt, egyre több szekvenciárészlet vált ismertté, illetve egy sor vírus teljes genomszekvenciáját is meghatározták. míg az első humán adenovírus genomszekvenálását (HAdV-2) 1986-ban, az első „atipikus mastadenovírusét (OAV287) és az első aviadenovírusét” (FAdV-1) csak 1996-ban, az első „atipikus aviadenovírusét” (TAdV-3) 1998-ban fejezték be, az első tipikus állati mastadenovírusét pedig (CAdV-1) 1997-ben. Jelenleg 21 adenovírus teljes DNS-szekvenciája ismert.

Az új adatok alapján a régi rendszer megváltoztatására tett korábbi javaslatokat ma már a genomszerkezetekben és génszekvenciákban fellelhető eltérésekre való hivatkozással is alá lehet támasztani. Az új javaslat szerint, melynek kidolgozásában laboratóriumunknak is fontos szerepe volt, négy nemzetséget különböztetünk meg. A két meglévő nemzetségből azokat a vírusokat, amelyek nem rendelkeznek E1A génnel a mastadenovírusoknál megtalálható E3 régióval, V és IX génekkel, de megtalálható bennük az ún. p32K szerkezeti fehérjét kódoló gén homológja, *Atadenovirus* néven új nemzetségbe soroljuk. A negyedik nemzetségbe egy, a leopárdbékából izolált adenovírus, a FrAdV-1 és az ezzel meglepő módon közeli rokonságot mutató, pulykából izolált TAdV-3 vírus tartoznak.

Az új rendszer kidolgozásánál laboratóriumunknak történeti okok miatt az *Atadenovirus* nemzetség leírásában volt igen fontos szerepe. A Rus szarvasmarha izolátum genomjának szekvenálásával célunk egyrészt az volt, hogy a javasolt új nemzetség egy újabb képviselőjéről gyűjtött információkkal bizonyítsuk, hogy e nemzetség leírt jellemzői ezen a rendszertani egységen belül általánosak, másrészt az újabb adatokkal bővítsük az adenovírusok törzsféjlődésének vizsgálatához szükséges adatbázist. Harmadrészt szándékunkban áll az atadenovírusok génkifejeződésének vizsgálata, amihez szükséges volt a kiválasztott modell szekvencia szintű ismeretére.

1.2. Módszerek

1.2.1. Molekuláris és mikrobiológiai technikák

Vírus-DNS tisztítása

A Moszkvai Mezőgazdasági Akadémiáról kapott „Rus” nevű szarvasmarha adenovírus törzset 10% foetális borjú savót tartalmazó MEM^a tápoldatban növesztett, primer borjú here szövettényezeten szaporítottuk.

^amodified Eagles medium

A vírus részecskék tisztítása a sejtkárosító hatás (CPE) maximumakor, a sejttenyészet háromszori fagyasztása–olvasztása és a sejttörmelék centrifugálással való eltávolítása után ultracentrifugálással történt. A DNS-t proteináz K enzimmel történő emésztést és fenolos extrakciót követően etil-alkohollal csaptuk ki.⁷

PCR reakciók

A polimeráz láncreakciók körülményei megegyeztek az enzimek gyártóinak leírásával. (A tapadási hőmérsékletek 2–4°C-kal a primerek olvadáspontja alatt voltak. A szekvenáláshoz sokszorozott fragmentumok esetében a Taq polimeráz által okozott esetleges másolási hibákat 4 független reakció termékének összekeverésével küszöböltük ki.

Molekuláris klónozás

A vírus-DNS fragmentumok klónozásához pBluescript SK és KS „Phagemid”-eket⁷ használtunk. A restriktív enzimekkel történő emésztésekhez a gyártók által javasolt mennyiségeket és előírt puffereket használtuk.

A restriktív enzimek hatástalanítása eleinte fenol-kloroformos extrakcióval történt, amit alkoholos kicsapás követett. Később áttértünk a QBIOSYSTEMS által forgalmazott „Miniprep Express Matrix TM” használatára.

1.2.2. Számítógépes analízis

Homológiai keresése

Homológiai keresésére általában az NCBI Blast adatbázis-kereső szolgáltatását és az intézeti szerveren saját adatbázisokkal szemben futó blast2 programot vettük igénybe. Hasonló célra használtuk a „MEME”^b rendszer „MAST”^c szolgáltatását is, a „Fred Hutchinson Cancer Research Center” „Block Maker”^d által képezett „homológia blokkok” segítségével.

A fehérjék általános tulajdonságainak vizsgálatára a PredictProtein szerveren^e elérhető programokat használtuk.

^bMultiple EM for Motif Elicitation

^cMotif Alignment & Search Tool <http://meme.sdsc.edu/meme>

^dhttp://blocks.fhcrc.org/blocks/blockmkr/make_blocks.html

^e<http://www.embl-heidelberg.de/predictprotein/predictprotein.html>

Illesztések

A homológ fehérjék illesztése az esetek egy részében a Clustal W program alkalmazásával történt, amihez esetenként a Clustal X^f grafikus felület segítségét vettük igénybe^g. Hasonló célra, valamint az illesztések szerkesztésére a Michele Clamp által készített „Jalview” programot használtuk^h.

Az illesztések publikálható formáját az „alscript” formázó-program segítségével hoztuk létre.

Filogenetikai számítások

A filogenetikai számításokhoz a PHYLIP programcsomag egyes alkalmazásait használtuk.

1.3. Az eredmények összefoglalása

Munkánk során meghatároztuk a Rus szarvasmarha adenovírus izolátum szinte teljes genomjának bázissorrendjét. A kapott szekvenciák számítógépes analízisével megállapítottuk a feltételezett fehérjék aminosav-szekvenciáit. A genomszerkezet és a kódolt fehérjék összehasonlító elemzésével bizonyítottuk, hogy a Rus törzs általános genomszerveződése eltér a *Mastadenovirus* nemzetségnél megszokottól, viszont nagyon hasonló a DAdV-1, OAV287 és BAdV-4 vírusoknál tapasztaltnak. Erre a vírusra is jellemző a többi adenovírus csoporttal homológ gének viszonylagos rövidege, a kódoló szakaszok közötti szokatlanul kis távolság, az E1B és E4 régiók eltérő összetétele, valamint az E1A, V, IX gének és a mastadenovírus-szerű E3 régió hiánya.

1. Az ITR szekvenciák összehasonlításából megállapítottuk, hogy azok összetétele és hossza a javasolt nemzetségekre jellemző eltéréseket mutat, amiből arra következtethetünk, hogy az adenovírusok DNS-replikációjáról főként a HAdV-2 és HAdV-5 vizsgálata alapján szerzett ismeretek egy jelentős része nem általánosítható.
2. Az LH korai régiót elemezve megállapítottuk, hogy az polimerázzal átfedően kódolt, korábban különböző adenovírusoknál leírt feltételezett fehérje valójában „műtermék”, az LH2 fehérje az összes atadenovirális

^fEz a program az „NCBI SOFTWARE DEVELOPEMENT TOOLKIT” része

^gncbi.nlm.nih.gov

^hReferencia hűján az elérhetőség: <http://www.ebi.ac.uk/michele/jalview/>

megfelelőit figyelembe véve nem áll leszármazási kapcsolatban a mastadenovirális E1B 19K fehérjével, az LH3 pedig távoli, de határozott rokonságot mutat az E1B 55K mastadenovírus fehérjével.

3. Megtaláltuk a szintén csak erre a nemzetségre jellemző ún. p32K protein génjét. A szekvenciahomológiák keresése során azonosítottunk egy bakteriális protein-családot, amely a p32K feltételezett funkciójához hasonló szerepet tölt be a bakteriális sporuláció során, és emellett bizonyos alacsony fokú hasonlóságot mutat szekvencia-szinten is.
4. E fehérje C-terminális konzervált részét kódoló DNS-szakaszra olyan, diagnosztikai célra alkalmasnak látszó PCR-módszert dolgoztunk ki, amelyekkel a birtokunkban lévő kérődző eredetű adenovírus mintákból sikerült megsokszoroznunk a kívánt DNS-darabot. A PCR termékek szekvenciája alapján az atadenovírusok közé soroltunk két, eddig molekuláris módszerrel még nem jellemzett adenovírus izolátumot is. Megmutattuk, hogy a kapott szekvenciák alkalmasak a vizsgált atadenovírusok csoportokba rendezésére, noha ez a csoportosítás nem szükségképpen esik egybe más géneken végzett filogenetikai célú vizsgálatok eredményeivel.
5. A hexon és a DNS-polimeráz fehérjéken végzett filogenetikai számításokkal megmutattuk, hogy az *Adenoviridae* családba tartozó vírusok négy, egymástól viszonylag távol álló csoportot alkotnak. Az a tény, hogy különböző fehérjék alapján végzett ilyen irányú vizsgálatok eredményei a nagyobb csoportok szintjén gyakorlatilag megegyeznek a genomszerkezeti eltérések alapján végezhető csoportosítással, alátámasztja a víruscsalád újrafelosztásának szükségességét.
6. Ugyanezekből a vizsgálatokból, valamint a p32K protein gén homológok C-terminális végeinek DNS szekvenciáin végzett filogenetikai számítások alapján megállapítottuk azt is, hogy az általunk vizsgált Rus törzs nem sorolható a BAdV-7 típusba, mint azt korábban leírták, hanem egy új vírustípus. Eddigi vizsgálataink alapján azt szűrhetjük le, hogy az ismertté vált atadenovírusok közül a BAdV-4 vírussal mutatja a legközelebbi rokonságot.
7. Az úgynevezett E3 régió már ismert szekvenciájának analízise során meghatároztuk a rajta potenciálisan kódolt három fehérje aminosav-sorrendjét. Az így kapott aminosav-szekvenciákból megállapítottuk, hogy a vizsgált proteinek egy, az atadenovírusokra általánosan jellemző,

fehérje-csoport képviselői. A nukleotid- és aminosav-szekvenciák elemzésével megmutattuk, hogy ezek a fehérjék a genomok jobb végén lezajlott duplikációk során alakultak ki. Az elemzések alapján felfedtük a Rus vírus e régióján tapasztalt duplikációk valószínű sorrendjét is.

1.4. Summary

The adenoviruses comprise one of the most important virus family of vertebrates, especially from scientific point of view. They were first reported as distinct pathogenic agents in 1953. The finding that certain human adenoviruses can induce tumors in newborn hamsters made them the model of tumorigenesis instantly. Many of the steps of their replication have been studied as models of certain intracellular processes. The research of their transcription regulation revealed the maturation processes of the eucaryotic mRNA. Their DNA replication, involving relatively few factors made them ideal for the study of DNA replication in general. The protein primed DNA-synthesis was first studied in adenovirus systems. Numerous research techniques first applied to map the genomic localization of adenoviral mRNA have been applied subsequently to many other pro- and eucaryotic systems. Transport of proteins from the cytoplasm to the nucleus, assembly of virions and of nucleic acid-protein complexes were studied in adenoviruses too. Nowadays, the research of the interaction of adenoviral proteins with each-other and with different transport mechanisms and life cycle regulator proteins of the host cells are important topics of the studies on adenoviruses. Their relatively harmless nature, easy replication and well understood replication processes made them important candidate vectors also in gene therapy research. Due to these promising characteristics, several differently designed generations of adenovirus based vectors were developed in the past decades.

However, although the scientific literature dealing with adenoviruses could fill libraries, most of the results were obtained from a few human adenoviruses, mostly from HAdV-2 and HAdV-5. Concerning animal adenoviruses, in most of the cases the researchers only serotyped the isolated strains, or in case of new host organisms electron micrographs were often made without further molecular studies, assuming that the virion similar to HAdVs covers a similar genome, with the same overall characteristics. If we look up the results of the sequencing projects, we can see, that the first complete genome sequence of a human adenovirus (HAdV-2) was determined in 1986, and the genome sequences of the first animal adenoviruses were published a decade later.

With the spread of automated DNA sequencers, the situation changed dramatically in the last 6 years. Several complete genome sequences of various animal adenoviruses are determined year after year, and the amount of partial sequence data is increasing rapidly too.

However, despite the fact that the occurrence of adenoviruses was observed in each vertebrate class, the official taxonomy of this virus family still recognises only two genera, (*Aviadenovirus* and *Mastadenovirus*).

A proposal for the new classification of *Adenoviridae* was made recently based on the different characteristics of the four distinct virus groups. The new classification of the family *Adenoviridae*, according to the proposal, divides the adenoviruses to four genera. Besides the two conventional genera the establishment of two new genera was proposed, *Atadenovirus* and *Siadenovirus*. Members of genus *Atadenovirus* lack the E1A, V, and IX genes, and the E3 region completely. These viruses have a characteristic structural protein named p32K near to the left end of the genome, a putative E3 analogue region on the right end of the genome and an altered E4 like section between the fiber gene and the „E3” region. Candidate atadenoviruses are the subgroup 2 bovine adenoviruses, the duck adenovirus 1 (DAdV-1) and the ovine adenovirus isolate OAV287. The members of the genus *Siadenovirus* are the turkey adenovirus 3 (TAdV-3) and the frog adenovirus 1 (FrAdV-1), which surprisingly share a lot of important characteristics: a sialidase-like gene near to the left hand end of the very short (26 kilobasepairs) genome, lack of E1A, and E3 homologues and an E4 region with one gene.

The genomic organization of Rus proved to be identical to that of OAV287 and BAdV-4, and very similar to DAdV-1. The base composition of the 28482 bp long sequence was also in the range of atadenoviruses (64.4% AT). Comparing the inverted terminal repeat (ITR) sequences of all the examined adenoviruses, we have concluded that the length and organization of the ITRs are characteristically different in the four proposed genera. Thus, the majority of the knowledge on the mode and mechanism of adenoviral DNA replication compiled mainly by studying HAdV types 2 and 5, might not necessarily be found universally true for animal adenoviruses, especially for those from birds or lower vertebrates.

The analysis of the early region LH (on the left hand end of the Rus genome) revealed that the putative protein the gene of which overlaps the DNA polymerase gene (encoded by the complementary strand) is indeed an artifact. Furthermore, the protein product of the LH2 gene is not phylogenetically related to the 19K protein of the mastadenoviral E1B, while LH3 exhibits a weak but sound homology with the E1B 55K protein.

We have been able to identify the gene of p32K, a precursor protein

uniquely found in the genome of all atadenoviruses studied so far. Attempts to find homologous proteins in the databases by conserved motif searches resulted in the recognition of bacterial protein group, i.e. the small acid soluble proteins (SASPs) that is involved in packaging the DNA into the bacterial spores. Since atadenoviruses do lack an important mastadenoviral core protein (V), hypothesizing functional relatedness between p32K and SAS proteins seems to be reasonable.

Degenerate oligonucleotide primers, designed on the basis of the amino acid alignment of all known p32K sequences, were synthesized and tested in PCR. The method seemed to be feasible for the genus specific detection of atadenoviruses.

Distance matrix analyses performed with the hexon and protease gene sequences showed clustering of members of the family Adenoviridae into four, clearly separated clusters. This separation, which corresponded to the four different types of genomic organization found in the members of the two conventional (Mast- and Aviadenovirus) and two proposed (At- and Siadenovirus) genera, further supported the need of reclassification of the family. Similar analysis on the p32K sequences of different bovine atadenoviral serotypes revealed that strain Rus is not an isolate of BAdV-7 (as thought and published earlier), but represents a new type. From the data at hand, we have concluded that BAdV-4 seems to be the closest relative of Rus.

We have predicted the amino acid sequence of three proteins encoded by the putative "E3" region of Rus. Analysis of the amino acid sequences showed that these proteins are the representatives of a protein group generally characteristic for atadenoviruses. By nucleotide analysis, we have demonstrated that these genes might have been evolved by duplications or reiterations at the right hand end of the genome. The likely temporal sequence of these duplications was also established.