B 3141

KIS RNS-EK PRIMÉR ÉS SZEKUNDÉR STRUKTÚRÁJÁNAK ELEMZÉSE SZÁMÍTÓGÉPPEL

Egyetemi doktori értekezés

Írta:

HEGYI HEDVIG

Készült

a

MTA Szegedi Biológiai Központ

Növényélettani Intézetében

1991

TARTALOMJEGYZÉK

1.	BEVEZETÉS	1
2.	IRODALMI ÁTTEKINTÉS	2
2.1.	Az RNS szerkezetére vonatkozó kezdet vizsgálatok	2
2.2.	Az U RNS-ek jellemzése	5
2.2.1.	Az U3 funkciója	6
2.2.2.	Az U3 RNS másodlagos szerkezete	7
2.3.1.	A nukleoplazmatikus U RNS-ek másodlagos szerkezete	8
2.3.2.	A nukleoplazmatikus U RNS-ek funkciója	10
2.4.	Az skRNS-eket kódoló gének szerkezete	12
2.4.1.	Az U1, U2 és U5 RNS génjei	12
2.4.2.	Növényi U skRNS pszeudogének	13
2.4.3.	Többszörösen kódolt paradicsom U1 gének	14
2.5.	A viroidok	14
2.5.1.	A viroidok szerkezete és működése	14
3.	MÓDSZEREK	19
3.1.	A rendszer	19
3.2.	A felhasznált szekvenciák	19
3.3.	Az RNS másodlagos szerkezetét meghatározó program	19
3.4.	A periodicitáskereső progam	22
3.5.	Az evolúciós programcsomag	24
3.5.1.	A távolsági mátrix módszerek	24
3.5.2.	Diszkrét karakteres módszerek	26
3.6.	A NUCALN program	28
3.7.	A MULTALIN program	28
3.8.	Az ESEE program	29
3.9.	A D3HOM program	29
3.10.	Az ALINCO program	30
4.	EREDMÉNYEK ÉS ÉRTÉKELÉSÜK	31
4.1.1.	A viroidok periodikus tulajdonságai	31
4.1.2.	A viroidok periodikus tulajdonságainak értelmezése	36
4.1.3.	Az ASBV 80 nukleotidos periodicitása	38
4.2.	A paradicsom U1 RNS gének evolúciója	38
4.2.1.	A határoló régiókból származtatott törzsfák	39
4.2.2.	A kódoló régiókból származtatott törzsfák	39
4.2.3.	A tranzíciók és transzverziók eltérő aránya az egyes régiókban	42
4.2.4.	A tranzíciók és transzverziók eltérő megoszlása a kódoló régiókban	43
4.3.1.	Az U3 RNS szekvenciák illesztése	47
4.3.2.	A többszörös illesztés információtartalma	49
4.3.3.	Az U3 RNS másodlagos szerkezetének analízise	49
4.3.4.	Az U3 RNS funkciója	56
5.	ÖSSZEFOGLALÁS	64
6.	IRODALOMJEGYZÉK	67
7.	KÖSZÖNETNYILVÁNÍTÁS	81

RÖVIDÍTÉSEK

skRNS =	sejtmagi kis ribonukleinsav
mRNS =	messenger ribonukleinsav
tRNS =	transzfer ribonukleinsav
rRNS =	riboszomális ribonukleinsav
ETS =	external transcribed spacer
ITS =	internal transcribed spacer

1. BEVEZETÉS

Napjainkban tanúi lehetünk mind a számítástechnika és tudomány, mind a molekuláris biológia hatalmas léptű fejlődésének.

E két terület számos ponton találkozik, mitöbb, valóságos interdiszciplinává fejlődött az elmúlt évtized során. Annál is inkább kézenfekvő ez a tény, ha meggondoljuk, milyen óriási mennyiségű információt halmozott fel már eddig is a molekuláris biológia, amelyet a szekvenálási technikák fejlődésével rohamléptekkel (exponenciális mértékben) növel tovább, nélkülözhetetlen eszközzé téve ezzel a kutatásban a számítógépet.

Jelen munkában e hatalmas információhalmaznak csak egy kis részét céloztuk meg számítógépes vizsgálatainkkal, nevezetesen: a sejtmagi kis RNS-eket (különös tekintettel azok növényi megfelelőire, mivel laboratóriumunkban azok molekuláris szintű kutatása folyik), amelyek fontos szerepet játszanak a sejtmagban keletkező, nagy molekulatömegű prekurzor (pre-mRNS és pre-rRNS) RNS molekulák érésében; és a viroidokat, amelyek molekulatömegük miatt ugyancsak a kis RNS-ek kategóriájába tartoznak, és enigmatikus működésük miatt széles körben kutatják őket.

Számítógépes vizsgálataink során e kis RNS-ek primér szekvenciájának és szekundér struktúrájának sajátosságait feltárva (részben meglévő, részben saját fejlesztésű algoritmusok és programok felhasználásával), az azokból levonható funkcionális összefüggésekre igyekeztünk következtetni.

2. IRODALMI ÁTTEKINTÉS

2.1. Az RNS szerkezetére vonatkozó kezdeti vizsgálatok

Az 1989-es kémiai Nobel-díjjal jutalmazott (Sidney Altman és Thomas R. Cech nevéhez fűződő) felfedezés - amely az RNS-ek katalitikus aktivitását mutatta ki, és így valószínűleg megoldotta az élet eredetét kutató tudósok tyúk-tojás problémáját (mi volt előbb - fehérjék vagy az őket kódoló nukleinsavak?) - új megvilágításba helyezte az RNS-ek kutatását.

Már száz éve annak, hogy a lipcsei egyetem hisztológia professzorának, Richard Altmannak először sikerült proteinmentes RNS-t izolálnia élesztőből, de 50 évnek kellett eltelnie ahhoz, hogy az RNS egyik fő funkcióját, a fehérjeszintézisben betöltött kulcsszerepét felismerjék. Az 1950-es években kezdték meg az RNS-ek szerkezetére vonatkozó kísérleteket. Az első kísérletek egyikében, melyet dohánymozaik vírus RNS-évél végeztek, kimutatták, hogy az RNS, szemben a kettős szálú DNS-sel, egyszálú molekula. Ezt követően (1958-ban) Hall és Doty megmérték RNS-oldatok UV-abszorpciójának hőmérsékletfüggését, és meglepetésükre karakterisztikus olvadási görbéket kaptak, vagyis az RNS a hőmérséklet emelésével a DNS-hez hasonlóan (bár nagyobb hőmérsékleti tartományban) struktúrális változáson megy keresztül.

Az első, termodinamikailag megalapozott módszert Tinoco és mtsai (1971) dolgozták ki az RNS másodlagos szerkezetének a meghatározására. Egy egyszálú RNS hurokká záródásához szükséges szabadenergiaváltozásra a

$$DG = -2.3*RT(B-1.5*\log(m-1))$$

formulát állították fel, ahol R és B konstansok, T abszolút hőmérséklet, m pedig a hurokban lévő nukleotidok száma.

Ezt követően több szerző publikált kísérleti eredményeket a különböző szerkezeti elemek által okozott szabadenergiaváltozás mérésére vonatkozólag (Delisi és Crothers, 1971; Gralla és Crothers, 1973a; Tinoco és mtsai, 1973; Gralla és Crothers, 1973b). Borer és mtsai (1974) 19, tökéletes bázispárosodásra képes oligoribonukleotid hipokromicitásának hőmérsékletfüggéséből azt találták, hogy a hélix stabilitása nemcsak a bázisösszetételtől, hanem a bázisok sorrendjétől, azaz szekvenciájától is függ. Kísérleteik alapján megalkották a "legközelebbi szomszéd" modellt (vagyis nem az egyes bázispárokra, hanem a legközelebbi, egymással kölcsönható bázispárszomszédokra állapították meg a különböző szabadenergia-csökkenés értékeket), amely mindmáig érvényesnek (és a későbbi komputeres programok szempontjából jól algoritmizálhatónak) bizonyult.

Az első komputeres programok, melyek egy adott RNS legvalószínűbb másodlagos szerkezetének prediktálására születtek, általában felsorolták az összes, egymással bázispárosodásra képes komplementer régiókat (a Watson-Crick bázispárok mellett megengedve a G-U párokat is), majd az összes kompatibilis régiók lehetséges permutációit véve, kiválasztották ezek közül a legvalószínűbb struktúrát (Pipas és McMahon, 1975; Korn és mtsai, 1977). Studnicka és mtsai (1978) ú.n. "komplex hiperstruktúrákat" generáltak, és megmutatták, hogy nincs szükség az összes lehetséges szerkezet létrehozására ahhoz, hogy a

legjobbat kiválasszák, de az ő módszerük is meglehetősen bonyolult és nehézkes volt.

A probléma megoldására az első dinamikus algoritmust Nussinov és Jacobson (1980) fejlesztették ki, akik az összes lehetséges bázispárra kiszámították, hogy az adott bázispár járulékát kivonva az egy lépéssel előbb generált struktúra szabadenergiájából, mekkora energiájú szerkezetet kapnak. Ezzel az eljárással egy N*N-es (ahol N a szekvencia hossza) mátrix minden elemét kiszámítva, a végső lépésben megkapták a legvalószínűbb struktúrához tartozó szabadenergia-csökkenést. Ezt az algoritmust fejlesztette tovább Zuker és Stiegler, akik nem a bázispárok, hanem a kölcsönható bázispárszomszédok által okozott szabadenergia-csökkenést vették figyelembe. (Erről az algoritmusról a Módszerekben még részletesebben is szó lesz.) Az energiaadatokat Salser (1977) összefoglalójából vették át.

Később több szerző közölt (Freier és mtsai, 1986 és Turner és mtsai, 1987) módosított energiaadatokat az egyes struktúraelemekre. Kierzek és mtsai (1986) kimutatták, hogy a legközelebbi szomszéd modell alapján 10%-os hibahatáron belül jósolható meg az RNS kettőshélix stabilitása, ha az csak Watson-Crick bázispárokat tartalmaz. Mivel az instabilitást okozó struktúraelemekre (dudorok, hajtűhurkok, elágazási hurkok, belső hurkok) nincsenek megbízható adatok, ezért több szerző publikált olyan másodlagos szerkezet-meghatározó módszert, amely nemcsak az energiaminimumhoz tartozó szerkezetet, hanem több, szuboptimális szerkezetet is meghatároz (Williams és Tinoco, 1986; Yamamoto és Yoshikura, 1985).

Az RNS térbeli szerkezetéről nagyon kevés ismeretünk van. Mivel nagy RNS molekulákat különösen nehéz kristályosítani (Turner és mtsai,

1987), néhány tRNS-től eltekintve, egyetlen RNS-nek sem ismerjük a kísérletileg meghatározott, háromdimenziós szerkezetét. Ebben az irányban kevés az elméleti eredmény is: csak az ún. pseudoknotting ("pszeudokötés")-nek a tRNS-szerű RNS molekulák térbeli szerkezetének kialakulásában játszott szerepét vizsgálták Pleij és mtsai (1985). Csak legújabban történtek kísérletek RNS-molekulák háromdimenziós szerkezetének komputeres modellezésére: Westhof és mtsai (1989) részletesen elemezték egy eubakteriális és egy eukarióta 5S RNS térbeli modelljét, amelybe sztereokémiai kényszereket és az RNS-eket alkotó bázisok és foszfátcsoportok elérhetőségére vonatkozó kísérleti adatokat is beépítettek.

2.2. Az U RNS-ek jellemzése

Mivel munkacsoportunk az RNS-ek köréből az ún. sejtmagi kis vagy másnéven U RNS-eket kutatja, a következőkben összefoglalom ezekről alkotott leglényegesebb ismereteinket.

Már a 60-as évek elejére ismertté vált, hogy az emlős (majd később általánosabban eukarióta) sejtek sejtmagvai a mRNS-eken, tRNS-eken és rRNS-eken kívül tartalmaznak egy kis molekulasúlyú, stabil, nagy uridintartalmú RNS csoportot is, melynek tagjait éppen ezért U-RNS-eknek nevezték el (Muramatsu és Busch, 1964; Hodnett és Busch, 1968).

A hat fő (U1-U6), a sejtmagban nagy kópiaszámban jelenlévő U-RNS-t már a 60-as évek végére leírták, részlegesen jellemezték (Weinberg és Penman, 1968), kimutatták, hogy in vivo proteinekkel komplexálva fordulnak elő a sejtmagban (Prestayko és Busch, 1968). Megállapították,

hogy ezek az skRNP-k (sejtmagi kis ribonukleoproteinek) prekurzor messenger és riboszomális RNS-ekhez kötődnek a sejtmagban (illetve az U3 a nukleoluszban), így már korán arra a következtetésre jutottak, hogy azok érésében játszanak szerepet (Sekeris és Niessing,1975; Calvet és Pederson, 1981).

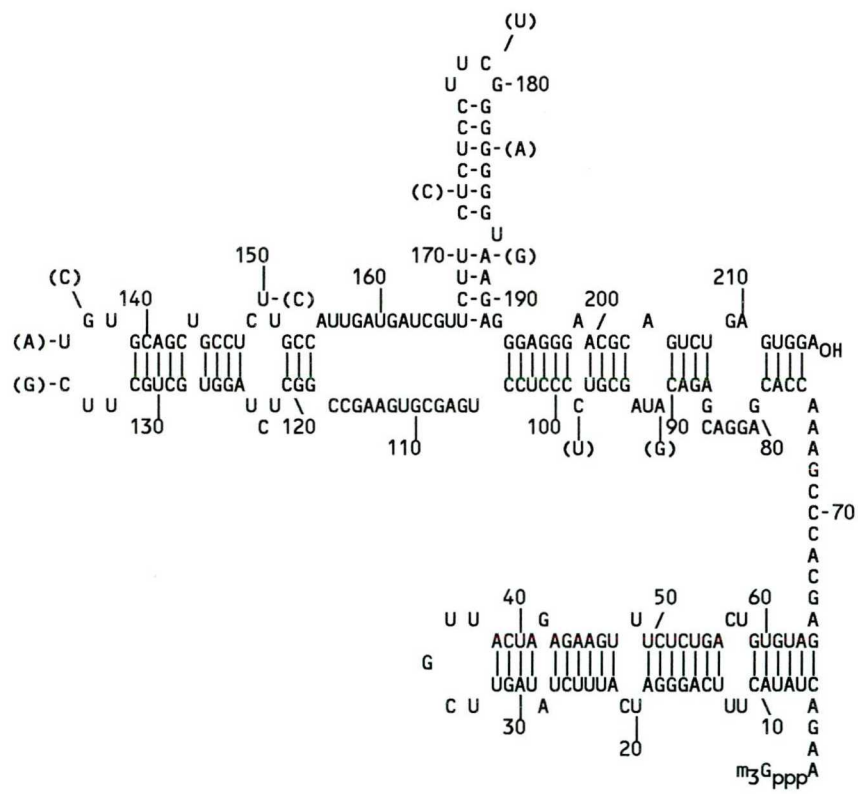
2.2.1. Az U3 funkciója

Az skRNS-ek kutatását jelentősen előremozdította ezen érési folyamatok részleteinek tisztázása. A pre-rRNS érésének egyes lépéseit már a 70-es évek elejére sikerült tisztázni. Mivel az U3 RNS (eltérően a többi, nukleoplazmatikus U RNS-től) a nukleoluszban található (Weinberg és Penman, 1968), feltételezték, hogy az U3 RNS a pre-rRNS érésében játszik szerepet (Prestayko és mtsai, 1970). Az U3 RNS-nek ebben a folyamatban való részvételére több modell jelent meg (Bachellerie és mtsai, 1983; Crouch és mtsai, 1983; Parker és Steitz, 1987). Bachellerie és mtsai szerint az U3 RNS bázispárosodással vesz részt az 5.8S rRNS és 28S rRNS közötti ITS2 szekvencia eltávolításában. E modellt később Tague és Gerbi (1984) módosították ITS2 szekvenciákból képzett konszenzus szekvencia alapján, mely szerint az U3 csak az ITS2 5' végi szekvenciájával bázispárosodna. A legújabb kísérletek viszont azt mutatták (Stroke és Weiner, 1989; Maser és Calvet, 1989), hogy az U3 RNS a riboszómális RNS prekurzorához, az RNS-től 5' irányban lévő ETS-hez (External Transcribed Spacer=külső átírt szakasz) kötődik, és ebből arra következtettek, hogy az U3 a riboszómális RNS "processing"-jének kezdeti lépésében vesz részt, mely az ETS 5' végének lehasadását eredményezi a

molekuláról.

2.2.2. Az U3 RNS másodlagos szerkezete

Patkány U3 RNS másodlagos szerkezetét kísérletesen vizsgálva, Parker és Steitz (1987) a 2.1. ábrán látható másodlagos szerkezetet kapták.



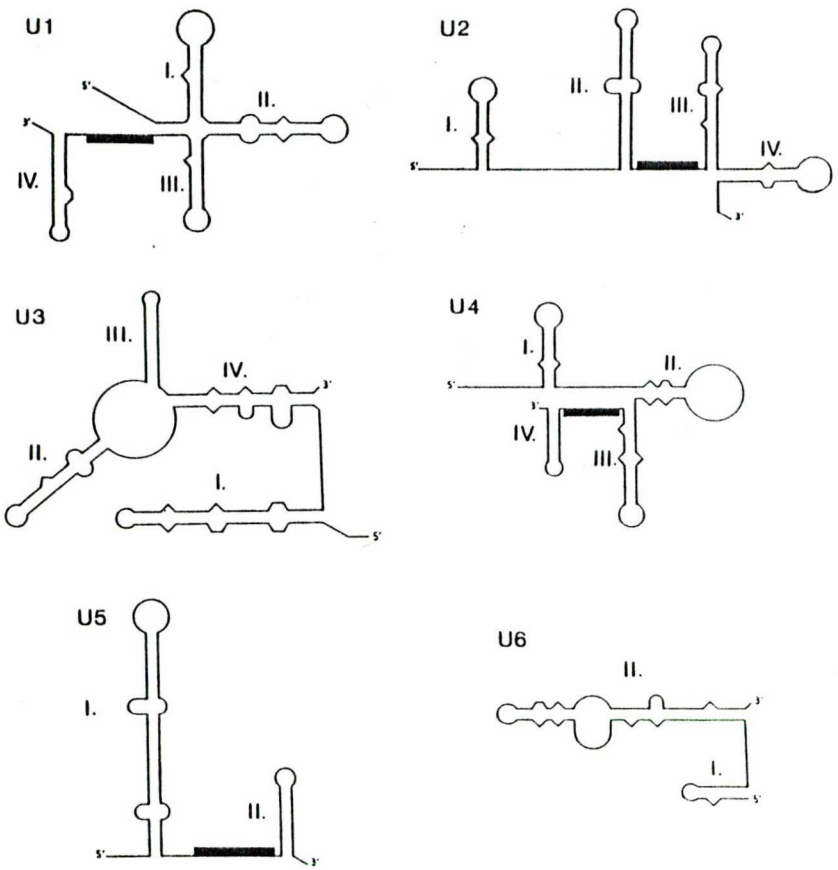
2.1. ábra: A patkány U3B RNS másodlagos szerkezete Parker és Steitz (1987) szerint. A zárójelben lévő nukleotidok az egér U3B szekvenciájának megfelelő változásokat mutatják.

Látható, hogy a molekula egy 5' végi rövidebb és egy 3' végi

hosszabb, erősen bázispárosodott törzsből (stem) és egy, a stemeket összekötő, egyszálú szakaszból áll. Ehhez hasonló másodlagos szerkezetet kaptak Jeppesen és mtsai (1988) *Xenopus laevis*ből izolált U3 RNS-re, viszont *Schizosaccharomyces pombe* U3 RNS vizsgálata alapján Porter és mtsai (1988) arra a következtetésre jutottak, hogy alacsonyabbrendű organizmusokból izolált U3 RNS-ek 5' végének másodlagos szerkezete jelentősen eltér a magasabbrendűekétől. E másodlagos szerkezetek érvényességét növényi U3 RNS szekvenciák ismeretében részletesen meg fogjuk vizsgálni.

2.3.1. A nukleoplazmatikus U RNS-ek másodlagos szerkezete

A patkány skRNS-ek másodlagos szerkezetét (az U3 kivételével) kísérletes és komputeres vizsgálatokkal Branlant és mtsai (1982) határozták meg, sematikus ábrájukat a 2.2. ábra mutatja. Látható, hogy az U1, U2, U4 és U5 molekulák közös strukturális elv alapján épülnek fel: mindegyikben található egy (az ábrán vastagon kihúzott) egyszálú régióban lévő $RA(U)_nGR$ konszenzus szekvencia, melyet kétoldalról két hajtűszerkezet (stem-and-loop) határol. E szerkezeti elemek együtt alkotják az ún. Sm fehérje-kötőhelyet. Az U6-ban nem található Sm-kötőhely, viszont *in vivo* bázispárosodással kötődik az U4-hez, amit kísérleti bizonyítékok mellett (Bringmann és mtsai, 1984; Hashimoto és Steitz, 1984; Rinke és mtsai, 1985) a molekulák koevolúciós sajátosságai is alátámasztanak. (A bázispárosodásban részt vevő szekvenciák az evolúció során mindkét molekulában úgy változnak, hogy a közöttük lévő



2.2. ábra: A hat fő U-RNS másodlagos szerkezete. A római számok a hajtűstruktúrák alkotta karokat jelölik. Az Sm-fehérje kötőhelyeket vastagon kihúztuk (Kiss, 1987).

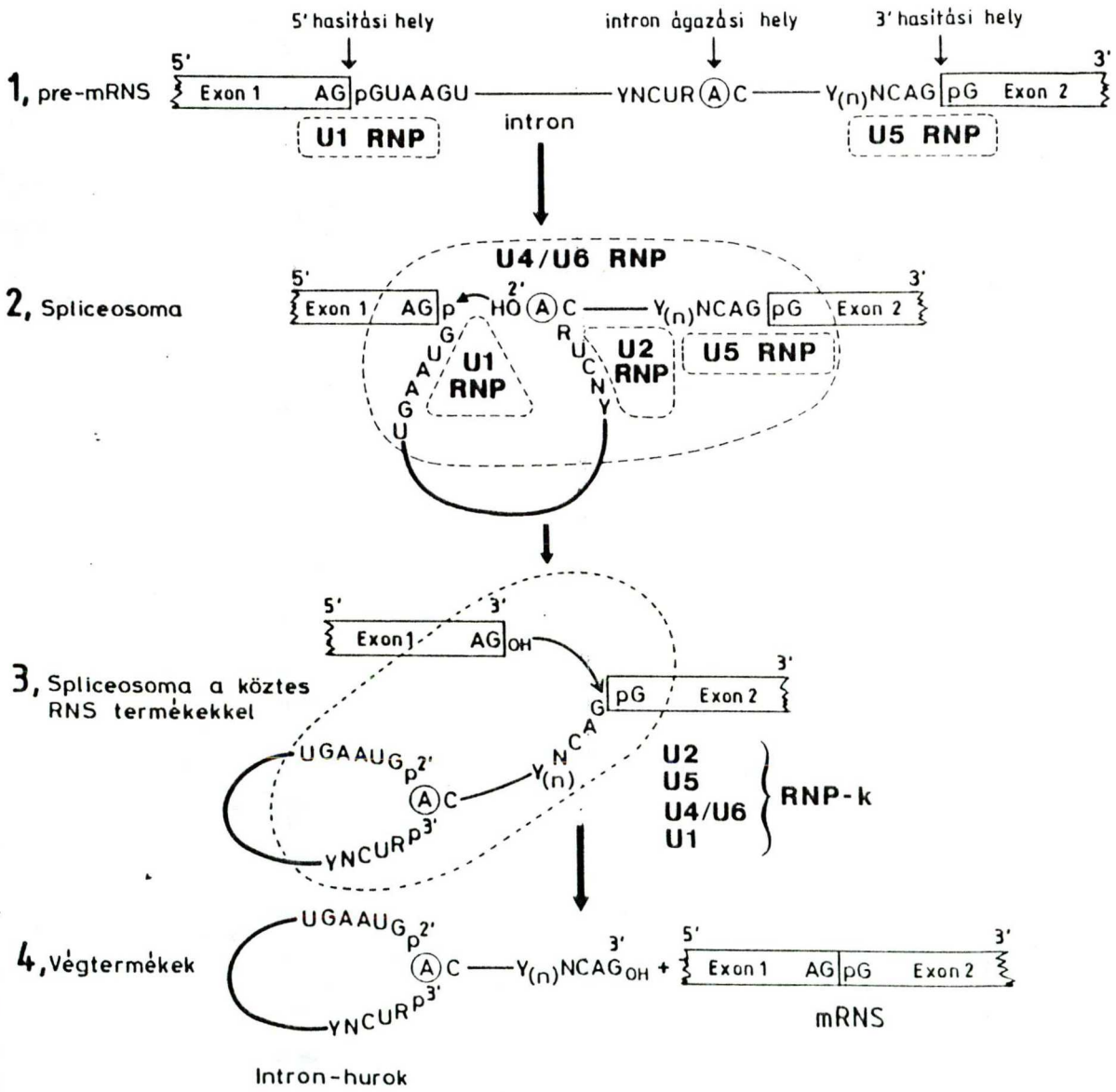
komplementaritás megmaradjon.) Így a modell érvényességét sikerült igazolni emlősökön kívül növényi (*Vicia faba*: Kiss és mtsai, 1988b) és élesztő (*Saccharomyces cerevisiae*: Guthrie és Patterson, 1988) U4/U6 szekvenciákra is.

2.3.2. A nukleoplazmatikus U-RNS-ek funkciója

Az eukarióta pre-mRNS érés egyik fontos lépése a vágás (splicing), amelynek során a nem kódoló, ún. intron régió kivágódik a pre-mRNS-ből, és a határoló, kódoló exon régiók újra összekapcsolódnak (Crick, 1979). Az U RNP komplexek a sejtmagban heterogén nukleoláris RNP (hnRNP) komplexekhez kapcsolódnak, (Brunel és mtsai, 1981), melyek a mRNS prekurzorokat tartalmazzák. Ez a tény arra engedett következtetni, hogy az U RNS-ek, illetve U RNP-k a pre-mRNS érésében játszanak szerepet. A folyamatról alkotott jelenlegi ismereteinket a 2.3. ábra foglalja össze.

Az eukarióta gének molekuláris szintű megismerése során felhalmozódó adatokból kitűnt, hogy az exon-intron régiókat határoló szekvenciák nagyfokú konzervativizmust mutatnak. E konzervált határoló szekvenciák (egész pontosan az intronok 5' végi hasítási helye) és az U1 RNS molekulák konzervált 5' végi szekvenciája között kimutatott komplementaritás alapján feltételezték (Lerner és mtsai, 1980), hogy az intronok és az U1 RNS-ek e konzervált régiói között bázispárosodás jön létre, ezáltal lehetővé téve a splicing első lépését, a pontos 5' végi vágást. E feltevést az a tény is alátámasztotta, hogy e kitüntetett 5' végi szekvencia az U1 RNS másodlagos szerkezetében egyesszálú régióban szabadon található (Rinke és mtsai, 1984). A modell helyességét Zhuang és Weiner (1986) kísérlete igazolta a legmeggyőzőbben, melynek során mutáns 5' végű intronok splicingjának hatékonysága *in vivo* megnőtt kompenzációs mutációt hordozó U1 RNS jelenlétében.

Harminckilenc, Drosophilából származó intron komputeres analízise alapján Keller és Noon (1985) egy olyan modellt javasoltak, amelyben az U2 RNS bizonyos régiói bázispárosodnak az intron-hurok



2.3. ábra: A pre-mRNS vágása. Az ábra az eukarióta pre-mRNS vágásáról alkotott mai ismereteinket összegzi. Részleteket lásd a szövegben (Kiss, 1987).

elágazási helyével és a 3' hasítási hellyel. A későbbi kísérletek megmutatták, hogy az U2 RNP valóban kötődik az intronhurok elágazási helyéhez, de magához a 3' hasítási helyhez nem (Black és mtsai, 1985; Chabot és mtsai, 1985).

Az U5 RNP jelenlegi feltételezések szerint az intron 3' vágási helyéhez kötődik, ennek pontos mechanizmusa azonban még nem ismert (Chabot és mtsai, 1985).

Az U4/U6 RNP-nek ugyancsak nélkülözhetetlen szerepe van a pre-mRNS vágási mechanizmusában, kötődési helye azonban nem ismert. Black és Steitz (1986) azt javasolták, hogy a vágási folyamat elején játszik szerepet.

2.4. Az skRNS-eket kódoló gének szerkezete

2.4.1. Az U1, U2 és U5 RNS génjei

A gerincesekben az skRNS géneket (U1-től U5-ig) a polimeráz II enzim írja át, ennek megfelelően 5' végi határoló régiójukban található egy proximális szekvencia elem (PSE) a -50 pozíció körül és egy disztális szekvencia elem (DSE) a -250 pozíció körül (Solymosy, 1990). A PSE funkcióját tekintve megfelel a TATA-boxnak (noha nincs explicit szekvencia hasonlóság), és a transzkripció korrekt iniciációjáért felelős. A DSE egy oktamer motívumot tartalmaz (ATGCAAAT), amely funkcióját tekintve egy enhancer-szerű elem. 3' végi határoló régiójukban a gének a +20 pozíció körül egy lazán definiált konzervált szekvenciát tartalmaznak, a "3' Boxot", ami a 3' vég pontos kialakulásáért felelős.

A növényekben az U1, U2 és U5 RNS-ek génjei 5' határoló régiójukban tartalmazzák a TATA boxot (konszenzus szekvenciája: TATAAA) a -30 pozíció körül, amely funkcióját tekintve megfelel a gerinces PSE-nek (Vankan és Filipowicz, 1989). Ezenkívül a -70 és -80 pozíció között tartalmaznak egy másik szekvencia elemet (USE = Upstream Sequence Element, konszenzus szekvenciája: RTCCCACATCG; Solymosy, 1990), amely azonban nem hasonlít a gerincesek DSE-éhez. Dohány protoplasztban végzett kísérletek szerint az USE mutációja vagy deléciója jelentősen lecsökkentette az U2 gén transzkripcióját. A 3' régió konzervált módon egy CA dinukleotiddal kezdődik, amelyet 4-9 nukleotiddal később egy konszenzus szekvencia követ (ARTNNAA), amely különbözik a gerincesek 3' Boxától.

2.4.2. Növényi U skRNS pszeudogének

Számos fajban találtak többszörösen kódolt U skRNS szekvenciákat. Ezek közül jelentős számú lókusz pszeudogénnek bizonyult (Dahlberg és Lund, 1988), amit főleg a humán genomban sikerült kimutatni. Jelentős mértékű multiplicitást találtak a növényi genomokban is: pl. U2-re és U5-re Arabidopsisban (Vankan és Filipowicz, 1988; Vankan és mtsai, 1988), vagy az U1-re babban (van Santen és Spritz, 1987), szójában (van Santen és mtsai, 1988), valamint laboratóriumunkban paradicsomban (Kiss és mtsai, 1988b).

2.4.3. Többszörösen kódolt paradicsom U1 gének

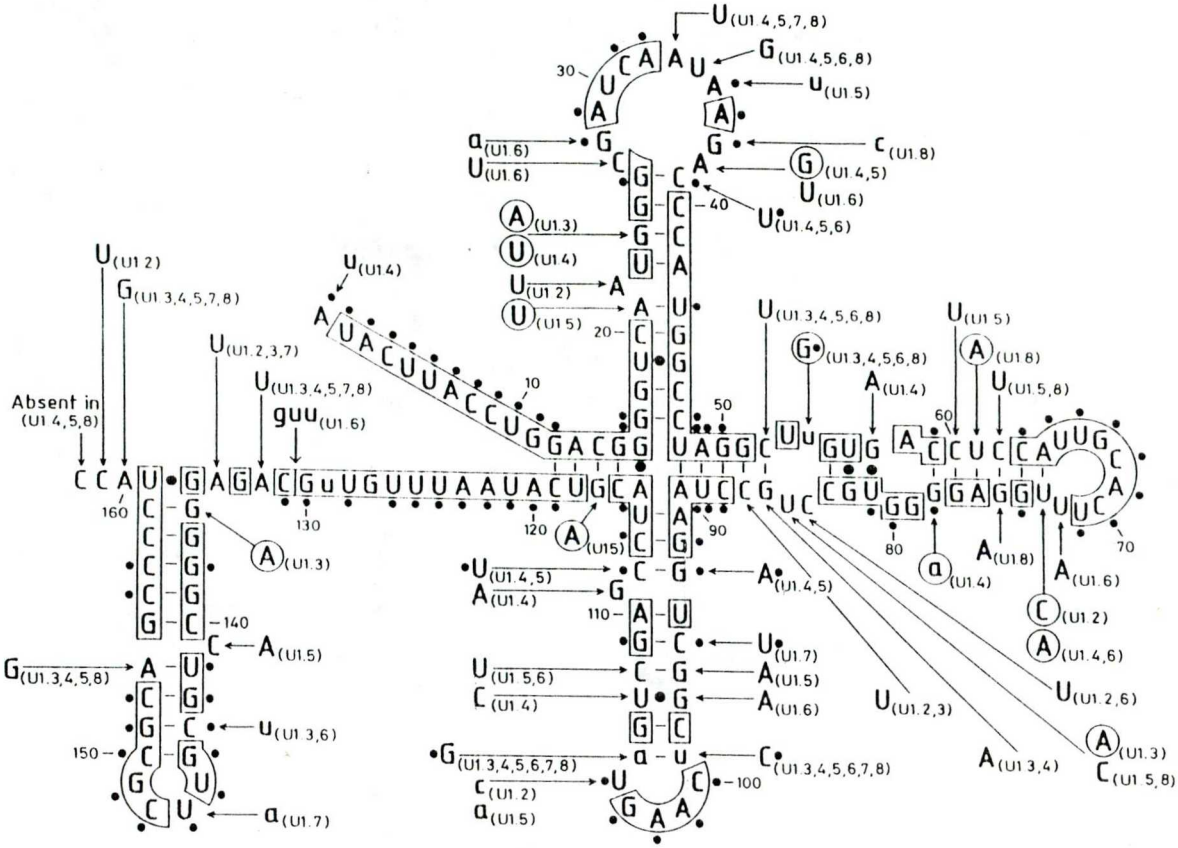
A levágott 3' végű U1 pszeudogénen kívül paradicsomból laboratóriumunkban 8 bona fide U1 gént is izoláltak, illetve megszekvenáltak (Abel és mtsai, 1989). A gének mindegyike tartalmazta a fent említett transzkripciós szignálokat (USE és TATA boxokat, illetve a 3' konszenzus szekvenciát), és a teljes kódoló régiókat. A gének közül az U1.1, az U1.2 és az U1.3, illetve az U1.4 és az U1.5 azonos lókuszban találhatók, tandem ismétlődésekben. A kódoló régiók közötti hasonlóság 85-96%-os, mindegyik variáns képes fölvenni az U1 RNS-re jellemző konszenzus másodlagos szerkezetet (Guthrie és Patterson, 1988; Kiss és mtsai, 1988a). Az U1.1 RNS másodlagos szerkezete (feltüntetve a többi variánsokban az U1.1 szekvenciájához képest bekövetkezett változásokat) a 2.4. ábrán látható. Az U1 variánsok szekvenciáit részletes komputeres (evolúciós) vizsgálatoknak vetettük alá.

2.5. A viroidok

Komputeres analízisünk másik fő tárgya a viroidok szekvenciája volt.

2.5.1. A viroidok szerkezete és működése

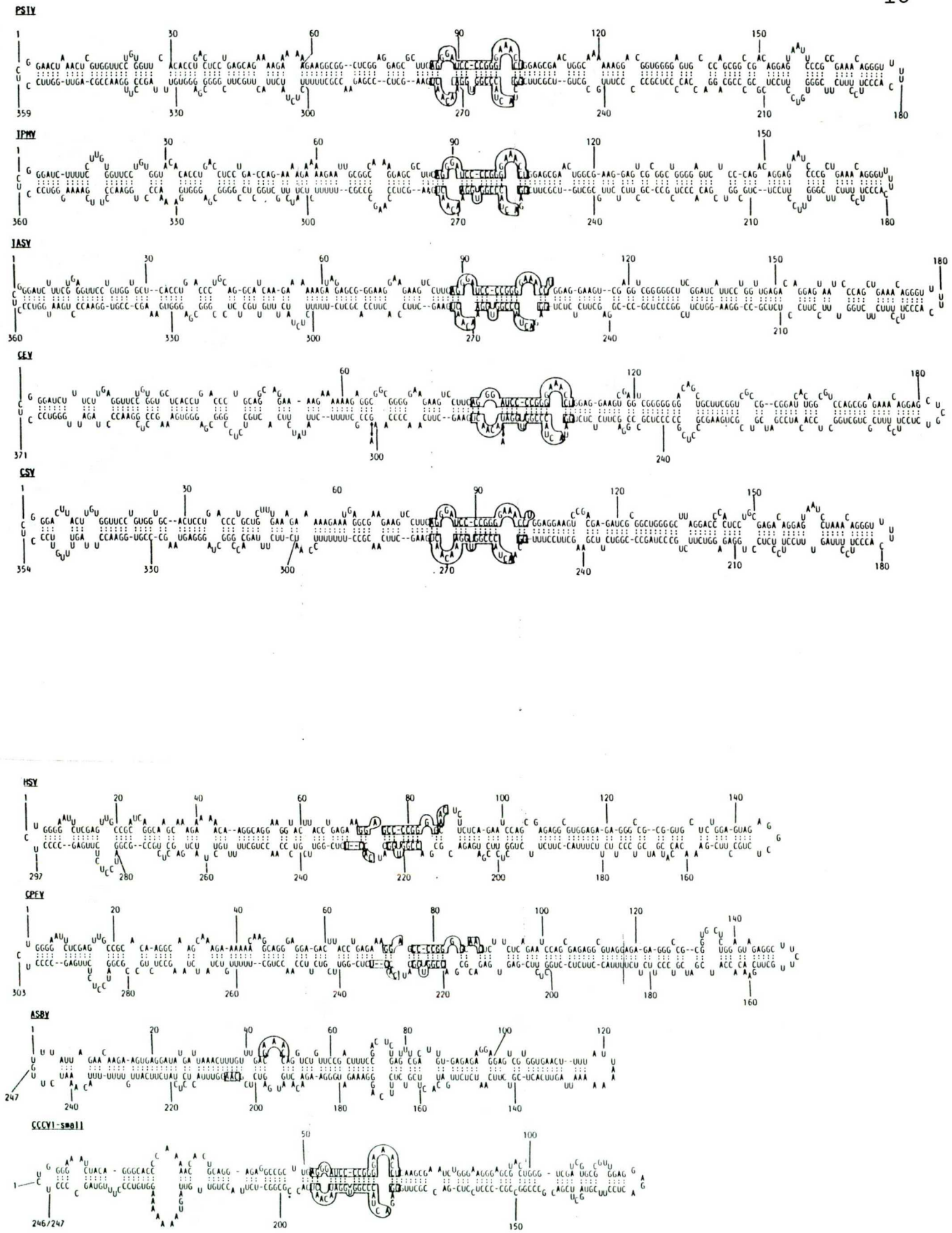
A viroidok kicsi (246-375 nukleotid hosszú), egyszálú, kovalensen zárt, cirkuláris RNS-molekulák, melyek növényi sejtekbe jutva patogén ágensekként viselkednek, és a gazdanövényben különböző



2.4. ábra: A paradicsom U1 RNS másodlagos szerkezete. Azokat a nukleotidokat, amelyek mind a nyolc U1 variánsban azonosak, bekereteztük. Az U1.1 szekvenciájához képesti változásokat nyilakkal jelöltük (feltüntetve a variáns nevét is, amelyben előfordul). Az U1-ben evolúciósan konzervált pozíciókat ponttal, a konszenzus szekvenciától (lásd Guthrie és Patterson, 1988) eltérő nukleotidokat kisbetűvel jelöltük.

betegségtüneteket produkálhatnak (cf. Riesner és Gross, 1985). Natív állapotban rúd alakú másodlagos szerkezetet vesznek föl, amit az energetikai számítások mellett (Riesner és mtsai, 1979; Nussinov és Jacobson, 1980) elektronmikroszkópos felvételek (Sogo és mtsai, 1973) is igazoltak. A 2.5. látható a legismertebb viroidok másodlagos szerkezete.

A vírusokhoz hasonlóan a viroidok is képesek önállóan szaporodni az arra fogékony növényi sejtekben, kis méretük és korlátozott genetikai



2.5. ábra: A legismertebb viroidok másodlagos szerkezete. A bekeretezések a centrális konzervált régiókat jelölik (Riesner és Gross, 1985).

információtartalmuk ellenére. A viroidok replikációja ellentétes polaritású RNS-intermedierek segítségével történik, és nem igényel DNS-t. A vírus RNS-ekkel éles ellentétben a viroidok nem kódolnak fehérjéket, ezért a viroidok replikációja nyilván a gazdanövény enzimjeinek segítségével történik (Branch és Robertson, 1984). Az (általában ellentétes polaritású) oligomer viroid formák jelenléte azt sugallja, hogy a replikáció egy ú.n. "gördülő kör" típusú mechanizmus segítségével történik (Branch és mtsai, 1981).

A viroidok patogén hatásának értelmezésére több modell is született. Schnölzer és mtsai (1985) feltételezései szerint a PSTV rúd alakú struktúrájának bal oldala, az ún. "virulencia moduláló" régió játszana szerepet egy még ismeretlen gazdafehérje megkötésében. Keese és Symons (1985) szekvenciahomológia alapján 5 strukturálisan és funkcionálisan is elkülönülő doménre osztotta föl a viroidok rúd alakú szerkezetét, náluk is a centrális konzervált régiótól (2.5. ábra) balra eső régió játszana szerepet a patogenicitásban.

Annak ellenére, hogy számos részletet sikerült tisztázni a viroidnak a gazdanövénnyel való kölcsönhatását illetően, a mintegy 10 féle különböző viroid (és jelentős számú variánsuk) működésére mindezidáig nem készült egységes modell. Világos, hogy egy ilyen kölcsönhatás esetén a viroidok által hordozott információ valmilyen módon kompatibilis a megtámadott sejt információs rendszerével. Mivel a viroidok nem kódolnak fehérjéket, ez az információ nem lehet a klasszikus genetikai kód, hanem valamilyen általánosabb módon a DNS struktúrája által meghatározott, mint például a DNS-csavar 10.5 nukleotidja vagy a nukleoszóma 80 nukleotidja által meghatározott strukturális információ.

A viroidok szekvenciájára vonatkozó vizsgálataink meg fogják

mutatni, hogy azok valóban hordoznak ilyenfajta 11-12 nukleotidos és 60, illetve 80 nukleotidos periódussal jellemezhető "második üzenetet".

3. MÓDSZEREK

3.1. A rendszer

Az összes program IBM PC-n, DOS operációs rendszer alatt futott, a saját fejlesztésű programokat TURBO PASCAL nyelven írtam.

3.2. A felhasznált szekvenciák

Az elemzett szekvenciák egy részét laboratóriumunkban szekvenálták, ezek helyben rendelkezésre álltak. Ilyenek voltak a paradicsom U1 gének (Abel és mtsai, 1989) és a paradicsom U3 gén (Kiss és Solymosy, 1990).

A szekvenciák egy másik részét az EMBL 19 nevű, CD-ROM-on tárolt szekvencia adatbankból vettük, a PCGENE nevű szekvencia-kezelő és analizáló program adatbank-kezelő funkciójának felhasználásával.

A szekvenciák harmadik részét (azokat, amelyek még nem szerepeltek az EMBL 19 adatbankban) számítógépes hálózaton keresztül kértük le közvetlenül az EMBL-től.

3.3. Az RNS másodlagos szerkezetét meghatározó program

A program algoritmusát Zuker és Stiegler (1981) írták le, mely a következőkben foglalható össze:

A szerzők először különböző struktúrákat definiáltak, ezek:

i/ törzs ('stem') - a kettősszálú, bázispárosodott régió,

ii/ hajtűhurok ('hairpin loop') - ha egy egyszálú RNS bázispárosodással ketősszálúvá záródik, a közrezárt egyszálú régió a hajtűhurok,

iii/ buborék hurok ('bulge loop') - egy kettősszálú törzsbe ('stem-be') egyik oldalon beékelődő, 'kidudorodó' nukleotid(ok),

iv/ belső hurok ('internal loop') - egy kettősszálú törzsbe mindkét oldalon beékelődő, egymással szemben lévő buborékok,

v/ elágazási hurok ('bifurcation loop') - olyan hurok, amiből két vagy több szomszédos stem indul ki.

Az 3.1. táblázatban találhatók a különböző struktúraelemekhez tartozó szabadenergia-változás értékek, melyek magukban foglalják egyrészt a PCFOLD program által ereditleg használt, Salser (1977) által összegyűjtött energiaadatokat, másrészt Freier és mtsai (1986) által meghatározott energiaadatokat, harmadrészt az általam módosított, (a program működése szempontjából adekvátabb) értékeket tartalmazza. A táblázatból látható, hogy egyedül a stem-et alkotó halmozott bázispárok ("stacking") okoznak szabadenergiacsökkenést, az összes többi elem destabilizálja a struktúrát.

Az algoritmus számbaveszi egy S_{ij} $(1<=i<j<=N)$ (egy N nukleotidból álló RNS molekula *i.* és *j.* nukleotidja által határolt) részszekvencia által fölvehető összes másodlagos szerkezetet, és kiválasztja közülük az optimálisat (a legkisebb szabadenergiájú állapotot). *i* értékét fokozatosan csökkentve, *j* értékét fokozatosan növelve, szukcesszív módon (az előzőleg optimalizált részstruktúrák figyelembe vételével) meg tudjuk határozni az összes S_{ij} részszekvenciára az optimális másodlagos szerkezetet, így az utolsó lépésben a teljes S_{1N} molekulára megkapjuk a

BASE PAIRING ENERGIES IN TENTHS OF A KCAL/MOLE
SALSER'S DATA
STACKING ENERGIES (UG = GU)

	GU	AU	UA	CG	GC
GU	-3	-3	-3	-13	-13
AU	-3	-12	-18	-21	-21
UA	-3	-18	-12	-21	-21
CG	-13	-21	-21	-48	-43
GC	-13	-21	-21	-30	-48

BULGE LOOP DESTABILIZING ENERGIES BY SIZE OF LOOP

1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
28	39	45	50	52	53	55	56	57	58	59	61	62	63	64	65	67

HAIRPIN LOOP DESTABILIZING ENERGIES BY SIZE OF LOOP

	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
CG CLOSING	999	999	84	59	41	43	45	46	48	49	50	52	53	54	55	57	59
AU CLOSING	999	999	80	75	69	64	66	68	69	70	71	73	74	75	76	77	79

INTERIOR LOOP DESTABILIZING ENERGIES BY SIZE OF LOOP

CLOSED BY	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
CG-CG	999	1	9	16	21	25	26	27	28	29	31	32	33	34	35	37	39
CG-AU	999	10	18	25	30	34	35	36	37	38	39	40	41	42	43	45	47
AU-AU	999	18	26	33	38	42	43	44	45	46	48	49	50	51	52	54	56

BASE PAIRING ENERGIES IN TENTHS OF A KCAL/MOLE
FREIER'S DATA
STACKING ENERGIES (UG = GU)

	GU	AU	UA	CG	GC
GU	-5	-5	-7	-15	-13
AU	-5	- 9	-11	-18	-23
UA	-7	- 9	- 9	-17	-21
CG	-19	-21	-23	-29	-34
GC	-15	-17	-18	-20	-29

BULGE LOOP DESTABILIZING ENERGIES BY SIZE OF LOOP

1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
33	52	60	67	74	82	91	100	105	110	118	125	130	136	140	150	158

HAIRPIN LOOP DESTABILIZING ENERGIES BY SIZE OF LOOP

	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
CG CLOSING	999	999	74	59	44	43	41	41	42	43	49	56	61	67	71	81	89
AU CLOSING	999	999	74	59	44	43	41	41	42	43	49	56	61	67	71	81	89

INTERIOR LOOP DESTABILIZING ENERGIES BY SIZE OF LOOP

CLOSED BY	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
CG-CG	999	8	13	17	21	25	26	28	31	36	44	51	56	62	66	76	84
CG-AU	999	8	13	17	21	25	26	28	31	36	44	51	56	62	66	76	84
AU-AU	999	8	13	17	21	25	26	28	31	36	44	51	56	62	66	76	84

BASE PAIRING ENERGIES IN TENTHS OF A KCAL/MOLE
FREIER'S DATA (MODIFIED BY H. HEGYI)
STACKING ENERGIES (UG = GU)

	GU	AU	UA	CG	GC
GU	-5	-6	-6	-15	-16
AU	-6	- 9	-11	-18	-23
UA	-6	- 9	- 9	-17	-21
CG	-16	-21	-23	-29	-34
GC	-15	-17	-18	-20	-29

BULGE LOOP DESTABILIZING ENERGIES BY SIZE OF LOOP

1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
33	52	60	67	74	82	91	100	105	110	118	125	130	136	140	150	158

HAIRPIN LOOP DESTABILIZING ENERGIES BY SIZE OF LOOP

	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
CG CLOSING	999	999	74	59	44	43	41	41	42	43	49	56	61	67	71	81	89
AU CLOSING	999	999	74	59	44	43	41	41	42	43	49	56	61	67	71	81	89

INTERIOR LOOP DESTABILIZING ENERGIES BY SIZE OF LOOP

CLOSED BY	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
CG-CG	999	8	13	17	21	25	26	28	31	36	44	51	56	62	66	76	84
CG-AU	999	8	13	17	21	25	26	28	31	36	44	51	56	62	66	76	84
AU-AU	999	8	13	17	21	25	26	28	31	36	44	51	56	62	66	76	84

3.1. táblázat. A Salser (1977), Freier (1987) és az általunk használt energiaértékek bázispárszomszédokra és az egyes destabilizáló szerkezeti elemekre.

szabadenergia-csökkenés értékét, és az ennek megfelelő szekundér struktúrát.

Az algoritmus előnye, hogy az összes lehetőséget számbaveszi, így ténylegesen a legkisebb energiájú szerkezetet határozza meg, hátránya viszont, hogy a memóriaigénye a szekvencia hosszával négyzetesen emelkedik.

Az algoritmusra PC-n fejlesztettünk egy programot, amely azonban a TURBO PASCAL 3.0 korlátozott lehetőségei (és az említett nagy tárigény) miatt maximálisan 170 nukleotid hosszú szekvenciára működött, így általában a program PCFOLD nevű, a szerzők által (FORTRAN környezetben) fejlesztett verzióját használtuk (amely maximálisan 425 nukleotid hosszú szekvenciára működött).

A szerzők különböző "kényszerek" lehetőségét is beépítették programjukba (kényszerített vagy tiltott bázispárosodás formájában - a felhasználó által meghatározott régiókra), amely lehetővé teszi kísérleti vagy szekvenciaösszehasonlításból származó információk figyelembe vételét.

3.4. A periodicitáskereső program

A szekvenciák belső hasonlóságainak és periodikus tulajdonságainak kimutatására egy egyszerű, TURBO PASCAL nyelvű programot fejlesztettünk ki:

Bevezetve a *k*-tuplett (*k* hosszúságú részszekvencia) fogalmát, az első lépésben az első *k*-tuplettet hasonlítottuk össze az összes többi, tőle 3' irányba eső *k*-tuplettel, előre megadott hasonlósági feltétel alapján (*k*

elemből m egyezést megkívánva, $m<=k$), purinok (R=A v. G) és pirimidinek (Y=C v. U) szempontjából.

A második lépésben a második *k*-tuplettet hasonlítottuk össze az összes, tőle 3' irányba eső *k*-tuplettel, stb.

Az utolsó, *(N-k)*-dik lépésben az *(N-k)*-dik *k*-tuplettet hasonlítottuk az *(N-k+1)*-dik (vagyis az utolsó) *k*-tupletthez.

Így, végsősoron az összes, *(N-k+1)* *k*-tuplettet összehasonlítottuk egymással.

Bevezettük a *d* változót, amellyel két *k*-tuplett (kezdőpontja) közötti távolságot jelöltük. Definiáltunk egy *f(d)* gyakoriságfüggvényt, amelynek *d* pontban felvett értéke azoknak az eseteknek a számát adja meg, amikor két, egymástól *d* távolságra lévő *k*-tuplett közötti hasonlóság legalább *m/k*.

Az *f(d)* gyakoriságfüggvényt egy átlagolási eljárással simítottuk, mely az $f_m(d)$ függvényt állította elő, a következőképpen:

$$f_m(d) = \sum_{i=d}^{d+l-1} f(i)/l.$$

Az egyes periódusokból, ahol lehetett, konszenzus szekvenciákat származtattunk, és azok információtartalmát is meghatároztuk a Shannon és Weaver (1949) által bevezetett, és Schneider (1986) által szekvenciákra alkalmazott "bizonytalansági" mértéket használva, a következőképpen:

Ha P_i jelöli az i-dik féle nukleotid gyakoriságát a konszenzus szekvencia bizonyos pozíciójában, akkor a bizonytalanság erre az *L* pozícióra:

$$H(L) = \sum_{i=1}^{4} P_i * \log_2 P_i \quad (bit/pozíció).$$

H maximális értéke 2 bit lehet, ez akkor áll fönn, ha mind a négyféle nukleotid egyforma gyakorisággal van jelen egy adott pozícióban.

Egy konszenzus szekvencia adott pozíciójában lévő információtartalmat a bizonytalanság csökkenésében kapjuk meg, vagyis:

$$I(L) = 2\text{-}H(L) \quad (bit/pozíció).$$

A konszenzus szekvenciában foglalt összinformációt a bizonytalanságban bekövetkező teljes csökkenés adja:

$$I_{consensus} = \sum_{L} (2\text{-}H(L)) \quad (bit).$$

3.5. Az evolúciós programcsomag

A paradicsom U1 gének evolúciós analízisét a PHYLIP (Phylogeny Inference Package = filogenezist származtató csomag) nevű programcsomag segítségével végeztük el. A csomag két alapvetően különböző megközelítésen alapuló programokat tartalmazott: az egyik típus az ún. távolsági mátrix módszereket foglalta magában, a másik a "legnagyobb takarékosság elve" alapján működő módszereket.

3.5.1. A távolsági mátrix módszerek

Ide tartoznak a FITCH és KITSCH programok. E programok bemeneti adataiként a vizsgált gének közötti páronkénti filogenetikai

távolságokat használjuk (amelyet általában úgy kapunk meg, hogy egy "alignment" program segítségével meghatározzuk a páronkénti százalékos hasonlóságokat, és a kapott értékeket kivonjuk 100-ból). Jelen esetben, mivel a páronkénti összehasonlításoknál a tranzíciók és transzverziók gyakorisága különbözött, és arányuk eltért a vizsgált gének kódoló és határoló régióira, a távolsági mátrix elemeinek kiszámítására a következő képleteket használtuk:

$$D_{ij(coding)} = Ti_{ij(coding)} + R_{(coding)} * Tv_{ij(coding)}$$
$$D_{ij(flanking)} = Ti_{ij(flanking)} + R_{(flanking)} * Tv_{ij(flanking)}$$

ahol

D_{ij}: a távolsági mátrix eleme,

Ti_{ij}: a tranzíciók száma i és j gén között,

Tv_{ij}: a transzverziók száma i és j gén között,

R: a tranzíciók és transzverziók páronkénti összehasonlításokból származtatott aránya,

mind a kódoló (coding), mind a határoló (flanking) régiókra.

E programok célja, hogy megtalálják azt a fát, amelynek esetében az

$$S = \sum_{i} \sum_{j} (D_{ij} - d_{ij})^2 / D_{ij}^P$$

összeg minimális lesz, ahol

D_{ij}: a megfigyelt távolság *i* és *j* gén között,

d_{ij}: a számított távolság *i* és *j* gén között, amelyet úgy kapunk, hogy összeadjuk a rekurzív módon számított fa szegmenseinek hosszát *i* és *j* gén

között.

P: a hatványkitevő, amely különbséget tesz az egyes módszerek között. A Fitch-Margoliash módszer esetében (amelyet a FITCH és KITSCH programok is használnak) *P* értéke 2.

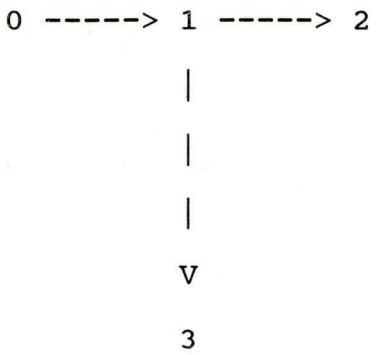
A FITCH és a KITSCH abban különböznek egymástól, hogy az utóbbi feltételezi egy evolúciós óra működését (ebben az esetben bármely gén a származtatott törzsfából számolva, a közös őstől azonos távolságra lesz), az előbbi nem. Mi ez utóbbi programot használtuk, mivel egy fajon belüli multigén család esetében az evolúciós óra megléte eléggé plauzibilis feltevés.

3.5.2. Diszkrét karakteres módszerek

Ebből a csoportból a MIX és a DNAPENNY programokat használtuk. Ezeknél a módszereknél azokban a pozíciókban lévő nukleotidokat kell megadni inputként a programnak, amelyekben különböznek egymástól a szekvenciák.

Sokal és Sneath (1963) talált egy módszert - melyet később Kluge és Farris (1969) fejlesztett tovább -, melynek a segítségével a többállapotú karaktereket (0,1) kétállapotú karaktersorozatokká alakítják át.

Így például, ha van négy állapot, amelynek a karakter-állapotfája így írható föl:



ahol 0 az ősi állapot, 1, 2 és 3 pedig a származtatott állapotok, akkor ez az új kétállapotú kódolással így írható föl:

régi állapot	új állapotok
0	000
1	001
2	011
3	101

A programok olyan törzsfát (vagy több, egymással ekvivalens törzsfát) generálnak, amely (v. amelyek) esetében a csomópontokban a báziscserék száma az egész törzsfát tekintve minimális (ezért hívják e módszereket "maximum parsimony"=legnagyobb takarékossági módszereknek is).

A távolsági mátrix módszerek esetében a program közös őssel ("gyökérrel") rendelkező fákat generál, míg a diszkrét karakteres módszereken alapuló programok gyökértelen fákat hoznak létre.

3.6. A NUCALN program

Két szekvencia gyors összehasonlítására a NUCALN (Wilbur és Lipman, 1983) programot használtuk, amely egy-egy értelmű hozzárendelést (ú.n. "alignmentet") hoz létre a két szekvencia megfelelő elemei között, olymódon, hogy a megadott paraméterek mellett a legtöbb találatot (így a legnagyobb homológiát) érje el a program. Az algoritmus a Needleman-Wunch-féle algoritmusnak (1970) egy továbbfejlesztett változata.

A program először számbaveszi az összes homológ k-tuplettet (k hosszúságú részszekvenciát) a két szekvencia által generált mátrixban, majd megkeresi e k-tuplettek közötti, optimális alignmenthez tartozó utat (egy előre definiált "ablakban", azaz ilyen szélességű sávban).

A program 3 féle, inputként megadható paramétert vesz figyelembe:

1. a k-tuplettek minimális hosszát,
2. az ablak szélességét,
3. a "gap penalty"-t, azaz a szekvenciaillesztés során létrejövő lyukra adott büntetést,

e paramétereknek általában függvénye az optimális alignment.

3.7. A MULTALIN program

Kettőnél több szekvencia illesztésére a MULTALIN (multiple alignment = többszörös illesztés) nevű programot (Corpet, 1988) használtuk. A program az első lépésben a megadott szekvenciák páronkénti

összehasonlítása alapján egy evolúciós fát generál, amit a további lépések során egy hierarchikus csoportosítás segítségével figyelembe vesz, olymódon, hogy először az evolúciósan közeli szekvenciákat illeszti, majd a kapott csoportokat illeszti egymáshoz.

3.8. Az ESEE program

Azokban az esetekben, amikor a MULTALIN nem adott jó eredményt vagy a szekvenciák olymértékben hasonlóak voltak, hogy nem volt szükség számítógépes illesztésre, az ESEE (Eyeball Sequence Editor="Szemgolyó Szekvencia Szerkesztő") nevű szerkesztő programot használtuk, ami tulajdonképpen nem más, mint egy szekvencia orientált szövegszerkesztő. A program lehetővé teszi több (maximum 21, mivel ekkora a teljes képernyőszélesség) szekvencia együttes kezelését, összehasonlítását és szemmel (ill. kézzel) való illesztését.

3.9. A D3HOM program

Hosszú szekvenciákban való lokális homológiák azonosítását a D3HOM program segítségével végeztük (Fristensky és mtsai, 1982), ami a mátrixos formában kapott eredményt tetszőleges léptékben képes összesűríteni, úgy, hogy a lényegi információ megmaradjon (A és Z közötti betűjellel jelöli a talált, 100 és 50% közé eső homológiát /A-val a 95-100%-os tartományba eső homológiát, B-vel a 90-94%-os tartományba esőt, stb.../ és a sűrítés során mindig az adott részmátrixban lévő legmagasabb homológiára utaló betűjelet őrzi meg).

3.10. Az ALINCO program

Az ALINCO saját fejlesztésű program egy többszörös illesztés információtartalmát (ALignment INformation COntent = Illesztés információ tartalma) számolja ki az előzőekben megadott Schneider-féle képlet alapján. Az *L* pozícióban bekövetkezett bizonytalanságcsökkenést (azaz információtartalmat) az

$$I(L) = 2 - \sum_{i=1}^{4} P_i * \log_2 P_i \quad (bit/pozíció)$$

képlettel kapjuk, ahol P_i jelöli az i-dik féle nukleotid gyakoriságát az illesztett szekvenciák *L* pozíciójában. A program inputjaként a többszörös illesztést kell megadni, outputként a pozíciónként kiszámolt információtartalmat ábrázoló görbét kapjuk.

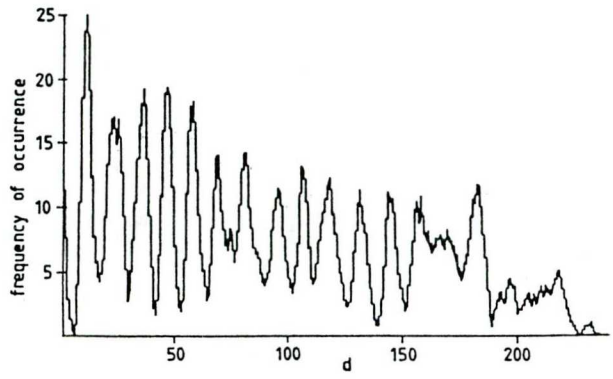
4. EREDMÉNYEK ÉS ÉRTÉKELÉSÜK

4.1.1. A viroidok periodikus tulajdonságai

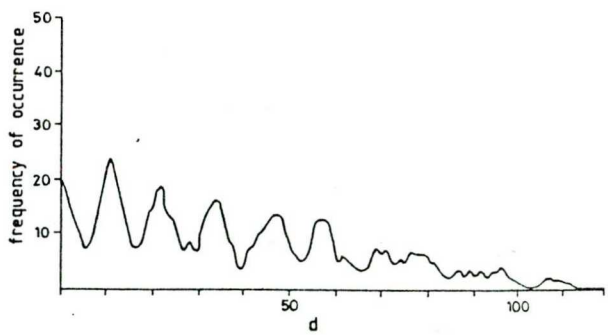
A Módszerekben leírt periodicitáskereső eljárást alkalmazva a legrövidebb ismert viroidra, a kókuszdió cadang-cadang viroidjára (coconut cadang-cadang viroid=CCCV), egy 12 nukleotid periódusú periodicitást találtunk, ami jól látható a 4.1. ábrán. Közelebbről megvizsgálva a 246 nukleotid hosszú CCCV szekvenciáját, sikerült megtalálni a molekula azon periodikus régióját, (a 244. és 129. nukleotid között), amely ezért a perodicitásért felelős (4.3. ábra). Egymás alá illesztve az egyes ismétlődési egységeket, egy CNGRRGRRAYCN konszenzus szekvenciát származtathatunk belőlük, mint az a 4.3. ábrán látható. E konszenzus szekvenciákkal szigorú egyezésben lévő szekvencia kétszer fordul elő a molekula felső felének ún. centrális konzervált régiójában (cf. Riesner és Gross, 1985) és egyszer a molekula alsó felében.

Ugyanilyen típusú, 12 nukleotid periódusú periodicitást találtunk (4.2. ábra) a PSTV (potato spindle tuber viroid) 60. és 180. nukleotidja által határolt régióban. Mint a 4.3. ábrán látható, ugyanaz - a CCCV-re jellemző - konszenzus szekvencia egymás után 9-szer fordul elő a PSTV-nek ebben a szegmensében.

Hasonló típusú periodicitást találtunk a PSTV csoport többi tagjában: a CEV (citrus exocortis viroid), a CSV (chrysanthemum stunt viroid), a TPMV (tomato planta macho viroid) és a TASV (tomato apical stunt viroid) viroidokban is.



4.1. ábra: A CCCV-re kapott $f_m(d)$ k-tuplett előfordulási gyakoriság függvény. A paraméterek: $k=6$, $m=5$, $l=4$ (a módszer leírását lsd. a 3.4. fejezetben).



4.2. ábra: A PSTV (60-180) régiójára kapott $f_m(d)$ k-tuplett előfordulási gyakoriság függvény. A paraméterek: $k=6$, $m=5$, $l=5$.

A

	C C U
	C U G G G G A A A U C -
	u A c A G G G c A C C C
CCCV1-small	C A a A A a c c A C u G
(244-129)	C A G G A G A G g C C G
	C U u G A G G G A U C C
	C C G G G G A A A C C U
	C A - A G c G A A U C U
	g G G A A G G G A g C G U A C
	C U G G G u c G A U C G
	u G c G c G - - - - - U
	u G G A G G A G A C u C
	C
consensus:	**C N G R R G R R A Y C N**

B

	C U
	C G G A G G A G - - C G
PSTV	C U u c A G G G A U C C
(69-172)	C C G G G G A A A C C U
	- G G A G c G A A C^{u}g G
	C A a A A a A G g a C G
	g U G G G G A G u g C C
	C A G c G G^{c}c G A C - -
	- A G G A G u A A U^{u}C C
	C - G c c G A A A C
consensus	**C N G R R G R R A Y C N**

4.3. ábra. A 12 nukleotidos ismétlődési egységek szekvenciájának illesztése és a belőlük konstruált konszenzus szekvenciák (A) a CCCV-re és (B) a PSTV-re. A centrális konzervált régiókat aláhúzással jelöltük.

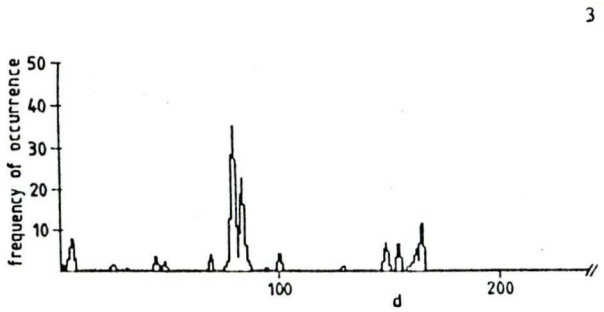
A HSV-ben (hop stunt viroid) és a CPFV-ben (cucumber pale fruit viroid) azonban, amelyek ugyanannak a viroidnak a variánsai, de csak távoli rokonságban vannak a PSTV csoporttal, a fenti konszenzus szekvencia ismételt előfordulása csak egy rövid régióra korlátozódik (a 69.-88. nukleotidok által határolt a HSV-ben), amely magában foglalja a centrális konzervált régió felső felét (ez utóbbi a HSV-ben a 73. és 88. nukleotidok között található).

Az ASBV (avocado sunblotch viroid), amely nem tartozik a PSTV csoportba, és nem tartalmazza a centrális konzervált régiót sem, a fentiektől eltérő típusú periodicitást mutat: a 247 nukleotid hosszúságú molekulában 80-as periodicitás mutatható ki (4.4. ábra), amely ilymódon 3 homológ részre osztja az ASBV szekvenciáját.

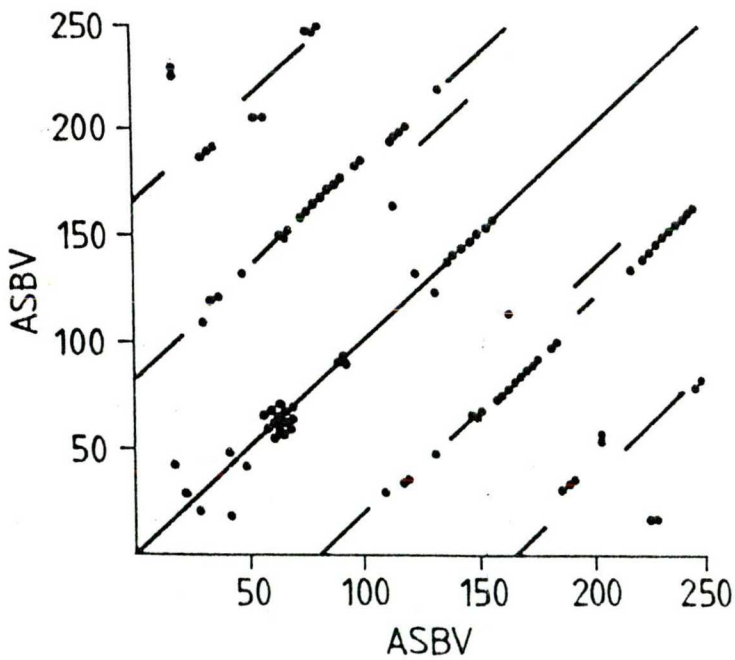
A 4.5. ábrán, amely az ASBV önmagával való összehasonlításának pontmátrixát mutatja, még szembetűnőbb a három periódus.

Egy később felfedezett viroid, az ASSV (apple scar skin viroid) (Hashimoto és mtsai, 1987), amely az eddigiektől külön csoportba tartozik, és ugyancsak nem tartalmazza a centrális konzervált régiót, egy harmadik típusú periodicitást mutat. A molekula szinte teljes egészében 60 nukleotid hosszúságú ismétlődési egységekből áll (4.6. ábra). Ezek 4.5-ször fordulnak elő a 330 nukleotid hosszú ASSV-ben .

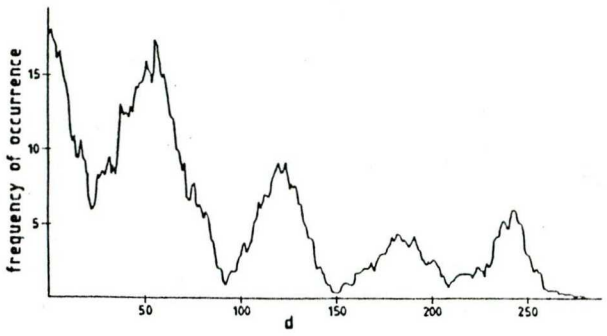
Az összes fönt említett periodicitások viroidspecifikusaknak tűnnek abban az értelemben, hogy nem találtunk ehhez hasonló periodikus sajátosságokat más, kis molekulatömegű RNS-ekben, így például sejtmagi kis RNS-ekben, virusoidokban, vagy olyan véletlen szekvenciákban, melyek hosszúságukat és nukleotid összetételüket tekintve megegyeztek a vizsgált viroidokkal.



4.4. ábra: Az ASBV-re kapott $f_m(d)$ k-tuplett előfordulási gyakoriság függvény. A paraméterek: $k=40$, $m=29$, $l=2$.



4.5. ábra: Az ASBV önmagával való összehasonlításának pontmátrixa. A paraméterek az előző ábráéhoz hasonlóak, vagyis 40 hosszúságú k-tuplettben 29 egyezést kívántunk meg.



4.6. ábra: Az ASSV-re kapott $f_m(d)$ k-tuplett előfordulási gyakoriság függvény. A paraméterek: $k=40$, $m=25$, $l=16$.

4.1.2. A viroidok periodikus tulajdonságainak értelmezése

Ismeretes, hogy a sejtekben lévő információtartalom megfejtését, a nukleinsavak által hordozott információk kiválogatását és dekódolását a nukleoprotein komplexek végzik.

Ennek a mechanizmusnak a kezdeti szakaszában (a nukleoszóma szinten és a transzkripció iniciáció szintjén) a felismerési eseményeket - melyek magukban foglalják a nukleoprotein komplexek kialakulását és szétesését - gyakorta a nukleinsavak periodikus tulajdonságai határozzák meg.

Ezt a jelenséget részletesen igazolták például a TF IIIA transzkripciós faktor és az 5S riboszomális RNS génje közötti kölcsönhatás esetében (Vincent, 1986). Egy másik bizonyíték a transzkripció iniciációs periodicitás létére az SV40 vírus enhancer régiója, amelyben lazán definiált motívumok ismétlődnek. Herr és Clarke (1986) elegáns mutációs

kísérletekkel igazolták, hogy az enhancer transzkripciós aktivitása akkor is helyreállt, ha e laza periodicitást az egyik motívum ismételt tandem duplikációjával helyettesítették.

E fenti példák analógiájára feltételeztük, hogy a PSTV-ben és a CCCV-ben és a velük rokon viroidokban talált 12 nukleotidos periodicitás fontos szerepet játszhat egy ilyen transzkripció iniciációs komplex-képző szabályozó fehérje hatásos megkötésében. A mintegy 10-szer egymás után megismétlődő szignál szekvenciák proteinkötő készsége olymértékben felerősödhet, hogy ezáltal a viroid sokkal hatékonyabban íródhat át, mint a sejt saját DNS-ei.

Az egymáshoz illesztett periódusok információtartalmát is meghatároztuk: ez a PSTV periodikus régiójára 7.3 bitnek, a CCCV megfelelő régiójára 6.7 bitnek adódott. Ezek az adatok jó egyezést mutatnak Schneider és mtsai (1986) által prokarióták genomjára végzett vizsgálatokból E.coli és Salmonella riboszóma-kötőhelyeinek információ-tartalmára kapott (8.3, illetve 10.6 bit) értékekkel.

Megjegyzendő, hogy a PSTV-t kísérletileg is proteinekkel (valószínűleg hisztonokkal) komplexálva találták a paradicsom nukleoluszának nukleoszomális frakciójában (Wolff és mtsai, 1985).

A strukturális periodicitás talált formái jó egyezést mutatnak a viroidok általánosan elfogadott filogenetikai csoportosításával. Bár a CCCV-t és a PSTV-t a korábbi osztályozások külön csoportba sorolták, eredményeink - miszerint a PSTV és a CCCV azonos, 12 nukleotid periódusú periodicitással rendelkezik, és a periódusokból származtatható konszenzus szekvencia a két viroidra teljesen azonos - összhangban vannak a legújabb filogenetikai vizsgálatokkal (Koltunow és Rezaian, 1989), melyek a CCCV-t több más viroiddal együtt besorolják a PSTV

csoportba.

4.1.3. Az ASBV 80 nukleotidos periodicitása

A sejtmaggal ellentétben az organellumokban nem ismeretes az RNS polimeráz polimorfizmus jelensége, és nincsenek polimeráz szelektív protein faktorok, ennek a ténynek egy lehetséges következménye, hogy a kloroplasztiszhoz asszociált ASBV (Mohamed és Thomas, 1980) nem tartalmaz ilyen, a DNS-csavar 10.5 nukleotidjával majdnem megegyező periódusú, 12 nukleotidos periodicitást.

A 80 nukleotidos periodicitás viszont, amely az ASBV-t jellemzi, releváns vonás az eukarióta kromatin (Weintraub, 1980) és a prokarióták nukleoid struktúrájában, és így szintén szerepet játszhat a nukleoprotein-komplex képződésében.

Az ASSV-ben talált 60 nukleotidos ismétlődési egységekkel jellemzett struktúrális periodicitás lehetséges szerepe egy esetleges nukleinsav-protein kölcsönhatásban jelenleg még nem ismert.

4.2. A paradicsom U1 RNS gének evolúciója

A laboratóriumunkban megszekvenált nyolc paradicsom "bona fide" gén (Abel és mtsai, 1989) kódoló és határoló régióiból különböző módszerekkel törzsfákat származtattunk.

4.2.1. A határoló régiókból származtatott törzsfák

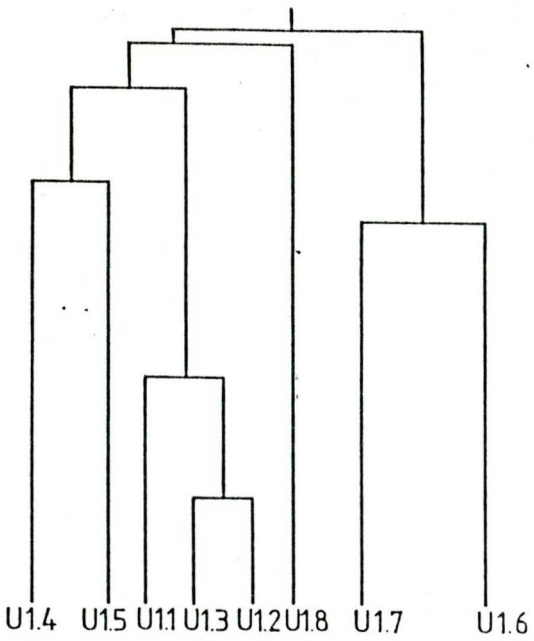
A 4.7.a ábrán látható az 5' és 3' végi szekvenciákat is tartalmazó határoló régiókból a távolsági mátrix módszerrel származtatott törzsfa, a 4.7.b ábrán pedig ugyanezekre a régiókra a legnagyobb takarékosság elvét alkalmazó módszerrel származtatott törzsfa.

Látható, hogy mindkét módszerrel származtatott törzsfában az U1.1-U1.2-U1.3 és az U1.4-U1.5 gének elkülönülő csoportokat ("clustereket") alkotnak, ami összhangban van a gének fizikai lokalizációjával a paradicsom genomban. Az U1.6, U1.7 és U1.8 helyzete azonban némileg eltér a két módszer esetében: a távolsági mátrix módszerrel kapott törzsfán ezen utóbbi gének elkülönülnek az U1.1-U1.2-U1.3 és az U1.4-U1.5 clusterektől, míg a diszkrét karakteres módszernél beékelődnek a két cluster közé.

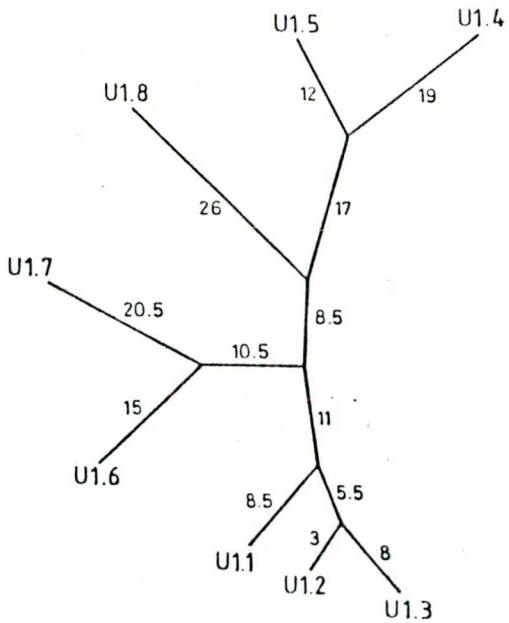
4.2.2. A kódoló régiókból származtatott törzsfák

A 4.8.a ábrán látható a kódoló régiókból a távolsági mátrix módszerrel, a 4.8.b ábrán pedig a diszkrét karakteres módszerekkel származtatott törzsfa. Látható, hogy bár az U1.5-U1.5 itt is külön clustert alkot, az U1.1-U1.2 és az U1.3 közé beékelődik az U1.7 génje. Ez utóbbi eredmény nincs összhangban sem a kísérleti adatokkal (az U1.7 külön klónban található, Abel és mtsai, 1988), sem a flanking régiókból származtatott törzsfákkal. Az U1.8 és az U1.6 helyzete némileg eltér a két módszer esetében: a távolsági mátrix módszer esetén az U1.8 van

A

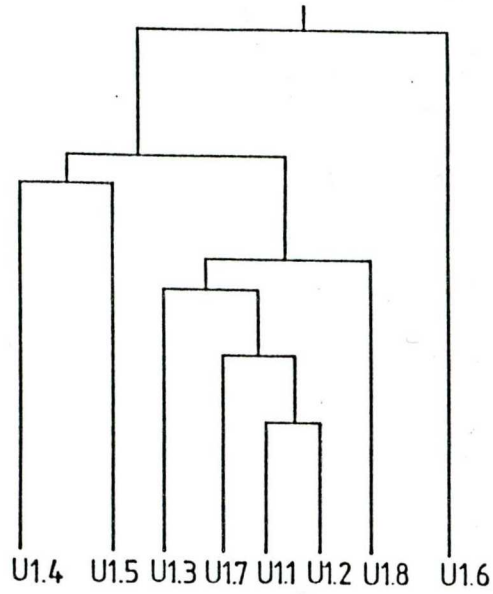


B

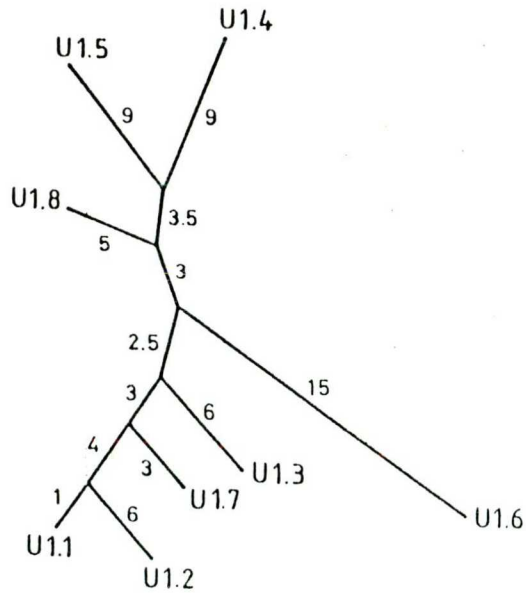


4.7. ábra: A határoló régiók szekvenciáiból származtatott törzsfák. (A) A távolsági mátrix módszerrel (KITSCH programmal) származtatott törzsfa. (B) Diszkrét karakteres módszerekkel (MIX és DNAPENNY programmal) származtatott törzsfa.

A



B



4.8. ábra: A kódoló régiók szekvenciáiból származtatott törzsfák. (A) A távolsági mátrix módszerrel (KITSCH programmal) származtatott törzsfa. (B) Diszkrét karakteres módszerekkel (MIX és DNAPENNY programmal) származtatott törzsfa.

közelebb az U1.1-U1.2-U1.3 clusterhez, míg a diszkrét karakteres törzsfában az U1.6.

A különböző módszerekkel, különböző régiókból származtatott törzsfák közötti különbségek értelmezésénél két fő tényezőt kell figyelembe venni:

i/ Az eltérő szelekciós nyomást, amelynek az egyes régiók feltehetőleg ki voltak téve. Valószínű, hogy az U1.7 eltérő viselkedése a kódoló és határoló (flanking) régiója tekintetében ennek a jelenségnek tudható be, a kódoló régiók ugyanis nyilvánvalóan erősebb szelekciós nyomásnak vannak kitéve, mint a határoló régiók, melyek - leszámítva a transzkripcióért felelős szignál szekvenciákat - funkció nélküliek. Elképzelhető, hogy az U1.7 génje korábban fizikailag is közelebb volt az U1.1 és az U1.2 génjéhez, és csak az utóbbi időben helyeződött át a genom egy másik lókuszába egy egyenlőtlen (unequal) crossing-over során.

ii/ A tranzíciók és transzverziók eltérő arányát a két régióban, ami az U1.6 és U1.8 gének helyzetének különbözőségére szolgálhat magyarázatul, mind a különböző módszerek, mind a különböző régiók vonatkozásában.

4.2.3. A tranzíciók és transzverziók eltérő aránya az egyes régiókban

A különböző módszerekkel generált törzsfák közötti különbségek értelmezésénél számottevő tényezőként kell figyelembe venni a tranzíciók és transzverziók eltérő gyakoriságát is. A nyolc paradicsom U1 gén páronkénti összehasonlításait figyelembe véve, azt kapjuk, hogy a tranzíció/transzverzió arány a kódoló régióra:

$$R_{coding} = 1.8158 \pm 0.22,$$

míg a határoló régióra ugyanez az arány:

$$R_{flanking} = 0.8051 \pm 0.11.$$

Behelyettesítve ezeket az arányokat a 3.4.1. pontban megadott képletekbe, az eltérő tranzíció/transzverzió arányokat a távolsági mátrix módszernél súlyozással figyelembe tudjuk venni, míg a diszkrét karakteres módszerek minden nukleotidcserét egyenlő súllyal vesznek figyelembe, ezért ott ilyen diszkriminációt nem tudtunk alkalmazni.

Az $R_{flanking}$ és az R_{coding} értékének ilyen mértékű eltérésének értelmezésénél plauzibilisnek látszott az a feltételezés, hogy a kódoló régióban előforduló majdnem kétszeres tranzíció/transzverzió aránynak a másodlagos szerkezet megőrzésében van szerepe, ugyanis minden C->U és minden A->G tranzíció esetében megőrződik a másodlagos szerkezet (csak C-G helyett U-G pár és A-U helyett G-U pár jön létre), míg a transzverziók esetében csak akkor őrződik meg, ha a stem szembenlévő pozíciójában is létrejön egyidejűleg a kompenzatórikus báziscsere.

Ennek a hipotézisnek az ellenőrzésére több pszeudogént is megvizsgáltam ebből a szempontból, például az ugyancsak laboratóriumunkban megszekvenált U3 pszeudogéneket (Kiss és Solymosy, 1990). Ezek a pszeudogének 5' trunkáltak, tandem helyezkednek el az igazi gén mögött, mindegyiküknek ugyanattól a ponttól hiányzik az 5' vége, tehát kialakulásukban minden bizonnyal DNS által közvetített események játszottak szerepet. A 4.9. ábráról leolvasható, hogy ennek ellenére sokkal

```
U3 GENE (97) GCCATGTGACCAGAGCGTGATTAACAGCTATCCATGGTTTCTCGAGCTGTGGTTA----CAGTAGA
U3.1ps (+54) ************************************************(+103)
U3.ps2 (+142)*************************************************----*******
U3.ps3 (+313)*********T*****************TC**T*****C*C**T***********G----*******
U3.ps4 (+478)*************************TC********C*C***A**T*******GGTTG*******


U3 GENE      TGATCGTTCTTAGACCCTTAATCTCAGGCCTAAGATGGTCTCATGGCTGTCTGACAGA (216)
U3.2ps      ******************-***T****A****************************** (+260)
U3.3ps      ************C*-***-G*****GA************************ (+424)
U3.4ps      **************-***-G******A*******************************A* (+599)
```

4.9. ábra. Az 5' végükön trunkált paradicsom U3 pszeudogének illesztése a valódi paradicsom U3 génhez. A szekvenciák Kiss és Solymosy (1990) munkájából valók.

nagyobb a tranzíciók aránya, mint a transzverzióké (14 tranzícióra csak 1 transzverzió jut).

Tehát nem feltétlenül az RNS másodlagos szerkezet megőrzésére való törekvés miatt áll fönn ez az arány, lehetséges, hogy ez valamilyen genomi sajátosság, mint ezt kimutatták főemlős mitokondriális gének vizsgálatánál (Brown és mtsai, 1982), ahol még sokkal nagyobb anomáliát tapasztaltak a tranzíció/transzverzió arányban (ez az arány ott 9:1-nek bizonyult), mint a paradicsom U1 gének esetében.

A nukleotidcserék szerkezeti megoszlásának azonban már szerepe lehet a stabil másodlagos szerkezet fönntartásában.

4.2.4. A tranzíciók és transzverziók eltérő megoszlása a kódoló régiókban

Véve az összes (7*8/2=28) páronkénti összehasonlítást, az összes tranzíciók száma, $N_{ti}=316$, míg az összes transzverzióké, $N_{tv}=180$.

T R A N S I T I O N S

TRANSVERSIONS	U1.1	U1.2	U1.3	U1.4	U1.5	U1.6	U1.7	U1.8
U1.1	------	3/1/1	8/2/0	11/2/0	13/2/0	9/0/1	4/0/0	8/0/0
U1.2	1/1/0	------	9/3/1	14/2/2	16/3/0	10/2/2	7/1/1	11/1/1
U1.3	0/3/1	1/2/1	------	7/3/1	10/4/0	9/4/1	6/2/0	5/2/0
U1.4	0/3/4	2/3/4	1/3/3	------	8/2/1	12/1/2	9/2/1	9/2/1
U1.5	1/1/6	2/2/6	1/2/3	1/2/4	-----	12/1/1	11/2/0	7/2/0
U1.6	1/3/1	3/3/2	1/4/2	0/3/3	1/3/5	-------	9/0/1	11/0/1
U1.7	0/1/3	1/0/3	0/2/2	0/4/3	1/2/5	1/4/4	------	6/0/0
U1.8	0/2/4	1/3/4	0/3/3	0/3/2	1/1/4	1/3/3	0/3/3	-----

In all:	Preserving	Changing	Neutral
TRANSITIONS	254	42	20
TRANSVERSIONS	22	69	89

4.1. táblázat. A tranzíciók és a transzverziók páronkénti megoszlása a paradicsom U1 kódoló régiókban (a másodlagos szerkezetet megőrző/megváltoztató/semleges értékek feltüntetésével). Alul a nukleotidcserék összesítése látható.

Az összes tranzíciók/transzverziók páronkénti megoszlását a másodlagos szerkezetben betöltött szerepét illetően, tehát, hogy megtartja, megváltoztatja a másodlagos szerkezetet vagy semleges arra nézve (vagyis egyesszálú régióban történik a nukleotidcsere), a 4.1. táblázat mutatja.

Összesítve a páronkénti összehasonlításokat, látható, hogy a tranzíciók többsége megőrzi a másodlagos szerkezetet, míg a transzverziók legnagyobb része a neutrális régióban található, így nem befolyásolják a másodlagos szerkezetet.

		TRANSITIONS								
		RU1	RUG1P	RUG1P1	RUG1P2	RUGP	HSU1	HRN03	HUG11P	HUG18P	MUG1P
TRANSVERSIONS	RU1	------	1/0/4	1/0/1	0/0/1	0/0/1	0/0/2	6/5/6	3/3/7	6/0/7	7/4/8
	RUG1P	0/1/1	------	2/0/5	0/0/5	1/0/5	1/0/6	7/4/8	4/3/6	7/0/11	8/4/9
	RUG1P1	0/0/0	0/1/1	------	1/0/0	1/0/2	1/0/3	7/5/6	4/3/8	7/0/8	9/4/9
	RUG1P2	0/0/0	0/1/1	0/0/0	------	0/0/2	0/0/3	6/5/6	3/3/8	6/0/8	7/4/7
	RUGP	0/0/0	0/1/1	0/0/0	0/0/0	-----	0/0/3	6/5/8	3/3/8	6/0/8	7/4/9
	HSU1	0/0/0	0/1/1	0/0/0	0/0/0	0/0/0	-------	6/5/5	3/3/5	6/0/4	7/4/8
	HRN03	0/3/1	0/3/2	0/3/1	0/3/1	0/3/1	0/3/1	------	5/7/9	6/4/11	4/9/13
	HUG11P	0/0/2	0/1/3	0/0/2	0/0/2	0/0/2	0/0/2	0/3/3	------	3/3/7	7/6/10
	HUG18P	0/1/0	0/2/1	0/1/0	0/1/0	0/1/0	0/1/0	0/2/1	0/1/2	------	9/4/13
	MUG1P	0/6/1	0/5/4	0/6/1	0/6/1	0/6/1	0/6/1	0/5/2	0/6/2	0/7/1	------

In all:	Preserving	Changing	Neutral
TRANSITIONS	184	104	283
TRANSVERSIONS	0	90	48

4.2. táblázat. Emlős U1 pszeudogének közötti tranzíciók és transzverziók páronkénti megoszlása (a másodlagos szerkezetet megőrző/megváltoztató/semleges értékek feltüntetésével). Alul a nukleotidcserék összesítése látható.

Kontrollként megnéztük az összes emlős U1 pszeudogéneket, hogy hogyan viselkednek ebből a szempontból (4.2. táblázat). Az összesítésből látható, hogy különbözik a megoszlásuk: jelentős számú tranzíció van, ami megváltoztatja a másodlagos szerkezetet, és sokkal több olyan, ami a neutrális, egyszálú régióban található, ugyanakkor nincs egyetlen transzverzió sem, ami megőrizné a másodlagos szerkezetet,

viszont nagyobbik részük megváltoztatja azt, és csak kisebb részük van a neutrális, egyszálú régióban.

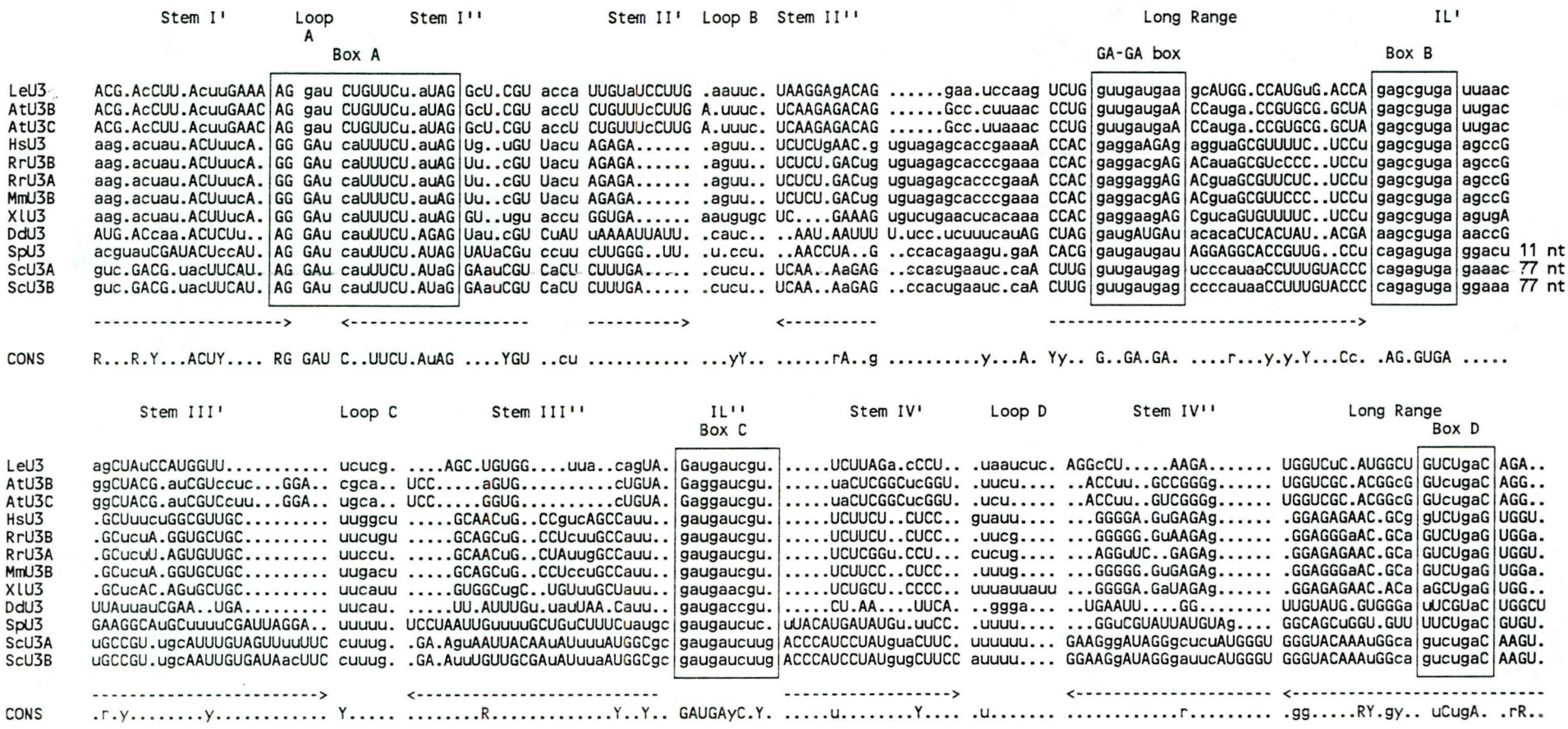
Tehát a paradicsom U1 gének közötti báziscserék megoszlása jelentősen különbözik az emlős U1 pszeudogénekétől, noha az utóbbiak esetében is jóval nagyobb a tranzíciók aránya, mint a transzverzióké, számszerűleg:

$$R_{ti/tv}=571/138=4.1377$$

Ez alátámasztja azt a máshonnan is valószínűsíthető feltevést, hogy legalábbis néhány paradicsom U1 gén esetében valódi, kifejeződő génről van szó. Ilyen valószínűsítő tényező, amellett, hogy az U1.1 gént Abel és mtsai (1988) ténylegesen is kifejezteték, hogy a többi géneknek is rendkívül hasonló a szerkezete, mindegyik variáns tartalmazza flanking régiójában a transzkripcióhoz szükséges szignálokat, és mindegyik fel tudja venni az U1-ekre jellemző konszenzus másodlagos szerkezetet, noha néhány esetben nem ez a szerkezet felel meg a szabadenergia minimumnak.

4.3.1. Az U3 RNS szekvenciák illesztése

A 4.10. ábrán látható az irodalomból elérhető szekvenciák többszörös illesztése ("multiple alignment"-je). A szekvenciákat részben a MULTALIN többszörös illesztőprogrammal illesztettük, részben pedig az ESEE szerkesztőprogrammal szerkesztettük az irodalomból ismert konzervált régiók (Box A,B,C: Wise és Weiner, 1980; Box D: Hughes



4.10. ábra: A teljes U3 szekvenciák többszörös illesztése. (Rövidítések: Le, paradicsom; At, *Arabidopsis thaliana*; Hs, humán; Rr, patkány; Mm, egér; Xl, *Xenopus laevis*; Dd, *Dictyostelium discoideum*; Sp, *Schizosaccharomyces pombe*; Sc, *Saccharomyces cerevisiae*.) Az egyes struktúraelemeket jelöltük az illesztés fölött, a konzervált boxokat bekereteztük. A *szekvenciákban* nagybetűvel jelöltük a bázispárosodott nukleotidokat, kicsivel az egyszálúakat. Az illesztés alatt lévő nyilak a stem-eket jelölik, az ez alatt lévő sor az illesztésből származtatott konszenzus szekvenciát tartalmazza. A *konszenzus szekvenciában* a nagybetű a teljes azonosságot, a kisbetű az ettől való egy eltérést jelenti.

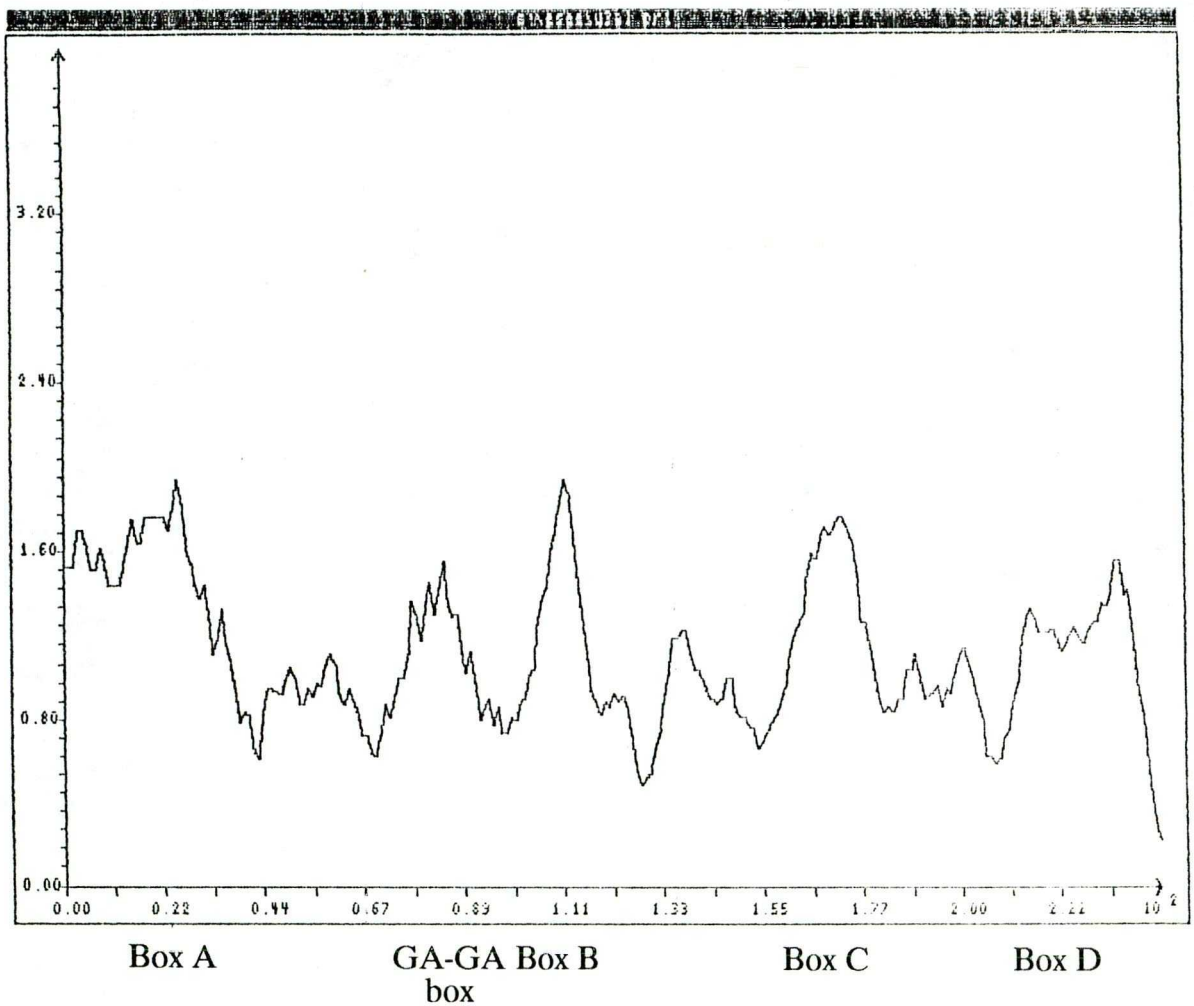
és mtsai, 1987; GA-GA Box: Porter és mtsai, 1988) egymáshoz illesztése és a másodlagos szerkezetek figyelembe vétele alapján, olymódon, hogy az egyes másodlagos szerkezeti elemek szekvenciái közötti legjobb illesztéseket kerestük meg. Ez a módszer sok olyan közös vonást is feltár az U3 RNS-ek szerkezetében, amelyek a szekvenciák közötti legnagyobb hasonlóságot kereső algoritmusok alapján nem nyilvánvalóak, de filogenetikailag és funkcionálisan realisztikusak.

4.3.2. A többszörös illesztés információtartalma

A 4.11. ábrán látható az illesztett U3 szekvenciákból származtatott információtartalom görbe, amely pozíciónként mutatja be a 4.10. ábrán látható illesztés információtartalmát. Látható, hogy mind az öt konzervált boxnak megfelel egy-egy helyi maximum a görbén, és az is leolvasható, hogy közülük az A box a legkonzerváltabb (a görbe alatti terület ennél a boxnál a legnagyobb).

4.3.3. Az U3 RNS másodlagos szerkezetének analízise

A paradicsom U3-ra kapott másodlagos szerkezet a 4.12. ábrán látható. A másodlagos szerkezet összes elemét a Zuker és Stiegler (1981) által fejlesztett PCFOLD nevű programmal határoztuk meg. Az ábráról leolvashatók az 5' végi stem-loopokra és a teljes szekundér struktúrára kapott energiaértékek. A számításokhoz szükséges energiaadatokat Freier és mtsai (1986) által a legközelebbi szomszéd modell alapján kapott

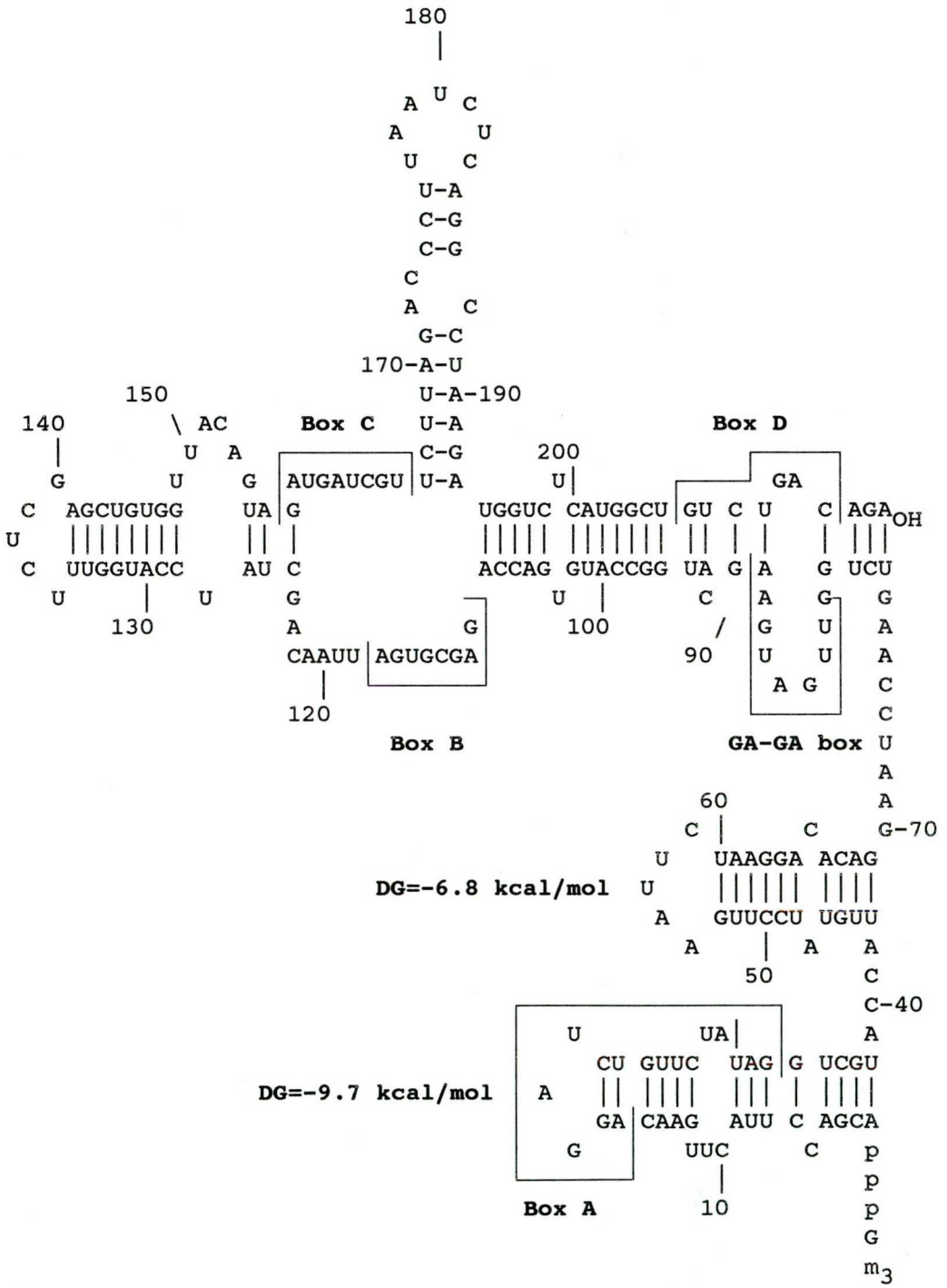


4.11. ábra: A 4.10. ábrán látható U3 RNS szekvenciák többszörös illesztésének információtartalma. A konzervált boxok helyét bejelöltük.

kísérleti értékekből nyertük. A program lehetőséget nyújt kísérleti adatokból nyert információk kényszerek formájában való beépítésére. A paradicsom U3 esetében két kényszert alkalmaztunk:

i/ az egyiket a központi stem 5'-3' végi bázispárosodására, mivel filogenetikai összehasonlításokból kitűnik, hogy az a régió mindig bázispárosodott, noha a primér szekvencia szintjén nem konzervált,

ii/ a másikat a B és a C box közötti bázispárosodás megakadályozására, mivel humán U3 RNS-sel végzett kémiai és



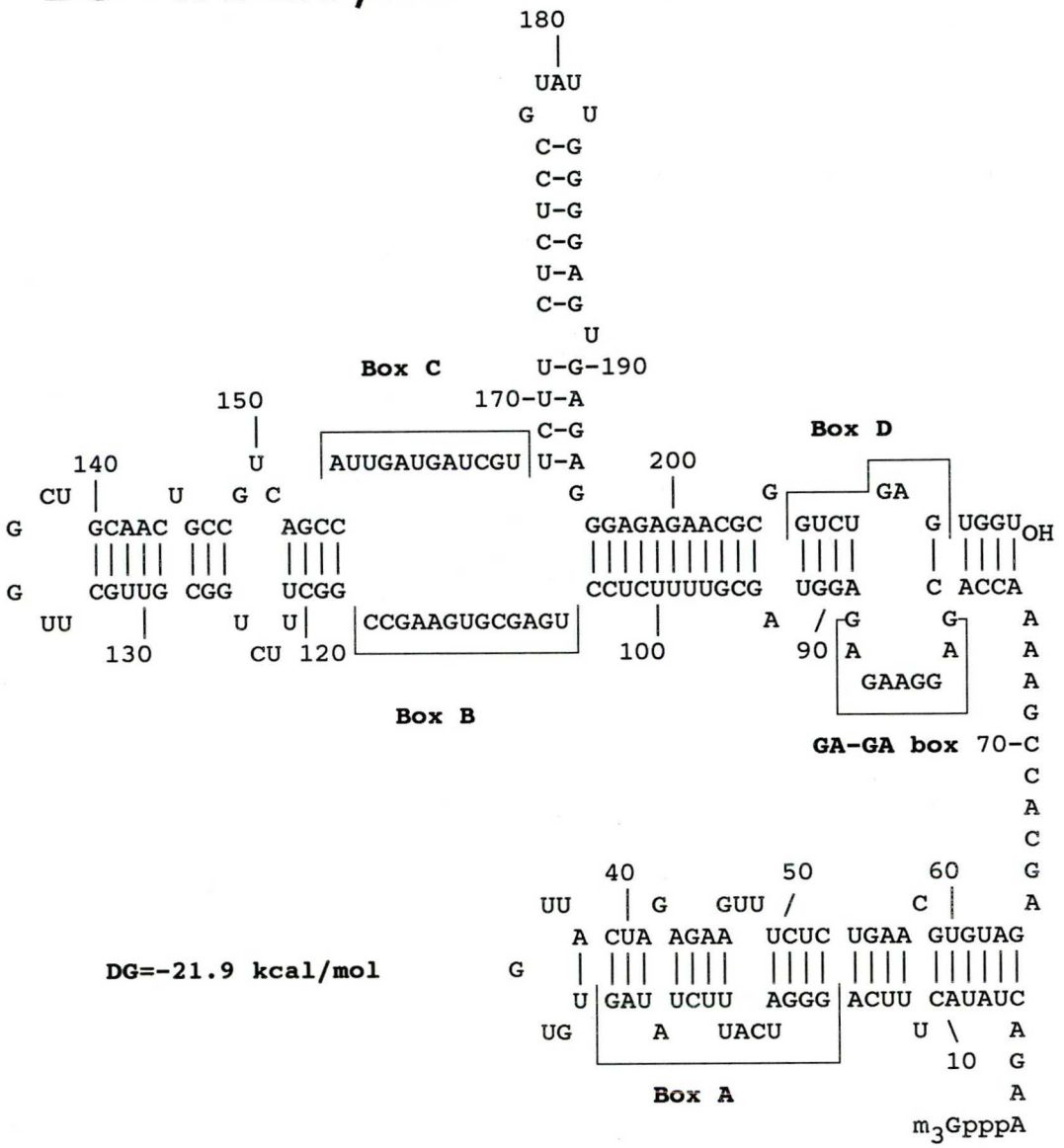
4.12. ábra: A paradicsom U3 másodlagos szerkezete a teljes struktúrára és az 5' végi stem-loopokra vonatkozó szabadenergia-csökkenés értékekkel.

enzimatikus próbákból Parker és Steitz (1987) úgy találták, hogy e két erősen konzervált régió egyesszálú, mitöbb, a fehérjékkel asszociált U3 RNP formát összevetve a "meztelen" U3 RNS-sel, azt találták, hogy e két régió az RNP formában még további védelmet nyer az RNS-hez képest az enzimatikus próbákkal szemben, megerősítve ezzel azt a feltételezést, hogy e két konzervált szekvencia fehérje-kötőhelyként funkcionál.

A GA-GA box és a D box a molekula központi törzsében helyezkednek el, egymással szemközt, jól definiált, invariáns pozícióban (szekvenciáikat tekintve is konzerváltan), egy belső hurkot alkotva.

A 4.13. ábrán látható a humán U3 RNS modell (Parker és Steitz, 1987) a szabadenergia-értékek feltüntetésével, ehhez nagyon hasonló másodlagos szerkezetet kaptak Jeppesen és mtsai (1988) a Xenopus U3 RNS-re. Látható, hogy a növényi U3 szekvenciája kísérletileg is ellenőrzött humán U3-éhoz nagyon hasonló másodlagos szerkezetbe rajzolható, ami a molekula 3' végét illeti, az 5' vége ezzel szemben eltérést mutat, amennyiben 1 helyett 2 stem-és-loopos szerkezetet képes csak fölvenni. A paradicsomon kívül növények közül ismert még az U3 teljes szekvenciája Arabidopsisból és részlegesen dohányból (Marshallsay és mtsai, 1990) (5' vég), valamint lóbabból (Kiss és mtsai, 1985) (3' vég). A növényi U3 szekvenciák átlagosan 70 %-os egyezést mutatnak egymással, így nem meglepő módon másodlagos szerkezeteik is nagyon hasonlóak, az ismert 5' végű növényi U3 RNS-ek mindegyike a 4.12. ábrán látható, a paradicsom U3-éhoz hasonló szerkezetet vesz föl.

A növényi U3 RNS másodlagos szerkezetéhez hasonló másodlagos szerkezetet kaptak (az 5' vég tekintetében is) Porter és mtsai (1988) *S. pombe* és más alacsonyabbrendű eukarióták U3 szekvenciáira. Ők ebből



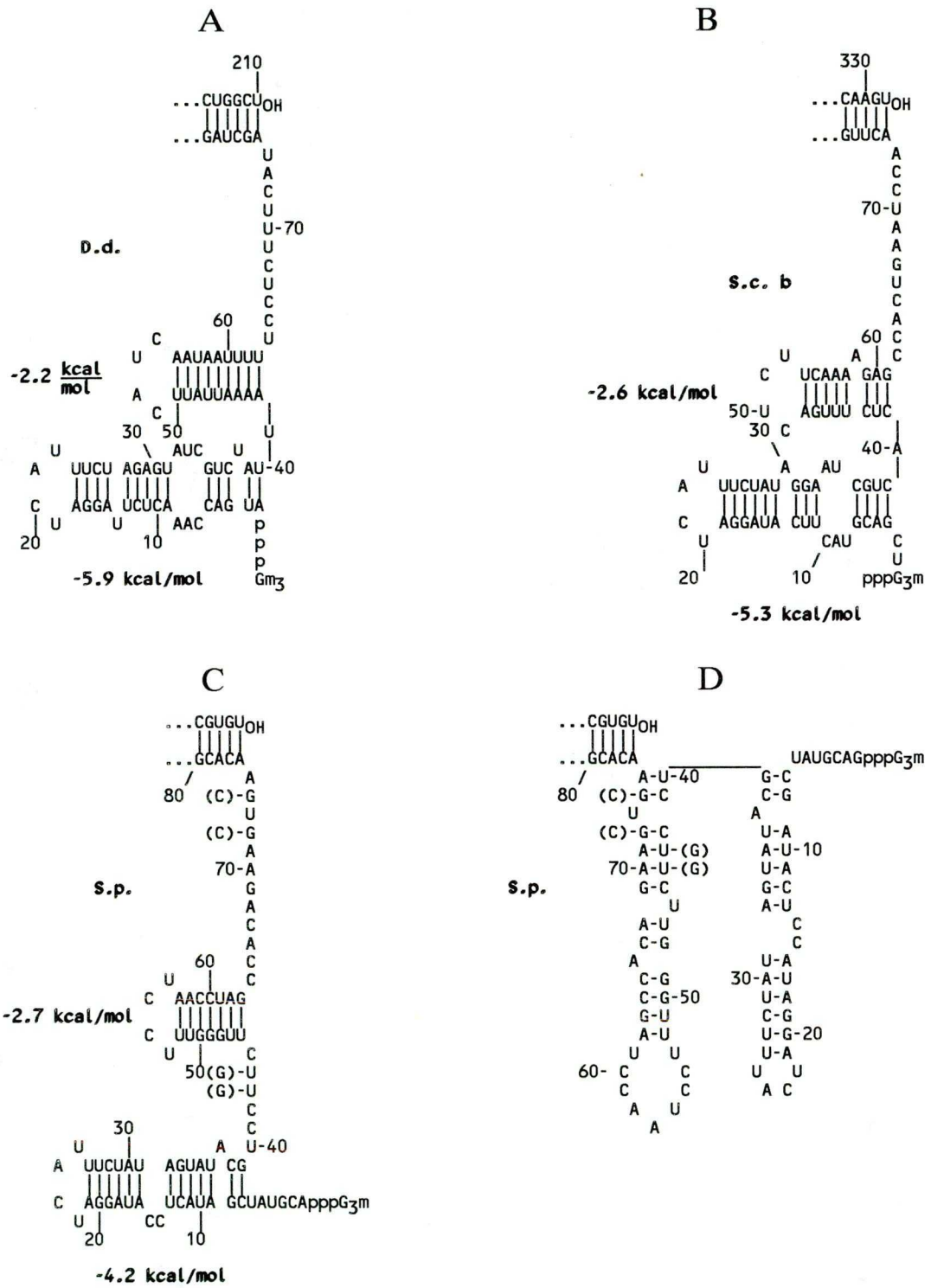
4.12. ábra. A humán U3 RNS másodlagos szerkezete a teljes struktúrára és az 5' végi stem-loopra vonatkozó szabadenergia-csökkenéssel.

azt a következtetést vonták le, hogy az U3 RNS-ek vonatkozásában a magasabbrendű és alacsonyabbrendű eukarióták között strukturális divergencia lép föl. Ellentmondani látszik azonban ennek a strukturális divergenciának az a tény, hogy ebben az 5' régióban található az egész U3 szekvenciában filogenetikailag legkonzerváltabb régió, az A box, amely a kétféle (egy és két hajtűhurkos) modellben így kétféle struktúrában

található (az egy hajtűhurkos modellben a stem-ben, míg a két hajtűhurkos modellben az első hajtűhurokban).

A 4.14. ábrán láthatók az alacsonyabbrendű eukarióták U3 szekvenciáinak 5' végeire kapott másodlagos szerkezetek. A Dictyostelium kivételével eltérnek ugyan a Porter és mtsai által kapott másodlagos szerkezetektől, viszont jobban hasonlítanak egymásra és a növényi U3 5' végi másodlagos szerkezetekre. Ez egyrészt a többszörös illesztésből látszik, ahol sikerült konszenzust kialakítani nemcsak a primér szekvencia szintjén, hanem a másodlagos szerkezet szempontjából is. Másrészt az alacsonyabbrendű eukarióták 5' végi hajtűinek energiaértékei jó egyezést mutatnak egymással (bár a növényi U3 energiaértékeknél [abszolút értékben] kissé alacsonyabb értékeket mutatnak). Harmadrészt, a mi modelljeink mindegyikében megtalálható egy egyesszálú régió, ami a molekula 5' végét és a központi törzset köti össze, amiről Parker és Steitz úgyszintén kimutatták, hogy erősen egyesszálú, (és így strukturálisan, azaz a molekula alakjának fenntartásában fontos szerepe lehet), míg Porter és mtsai *S. pombe* U3 modelljében ez a régió teljesen hiányzik. És végül, modelljeink érvényességét (*versus* Porter) az az utóbi időben megszekvenált szekvenciavariáns is alátámasztja, melyben az eredeti *S.pombe* U3 szekvenciájához képest bekövetkezett változások Porter és mtsai (1988) modelljében bázispárosodott régióra esnek, melyek azonban nem támasztják alá másodlagos szerkezet modelljüket (4.14.c ábra), mivel nem kompenzatórikusak, míg a mi modellünkben e nukleotidok egyesszálú régióban találhatók.

Utóbbi szerzők még egy lehetséges megoldást javasolnak az 5' vég primér szekvenciájának konzerváltsága és másodlagos szerkezetének heterogenitása közötti ellentmondás feloldására: hogy nincs egyáltalán



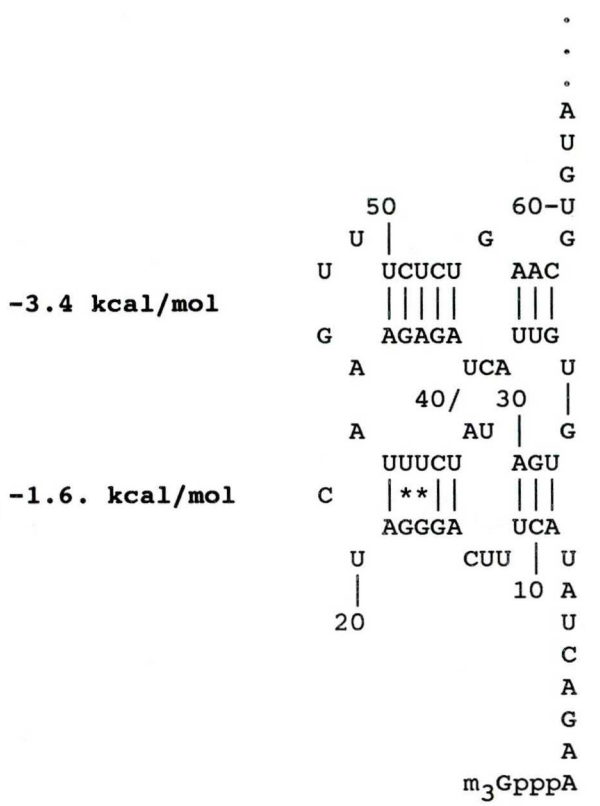
4.14. ábra: Az alacsonyabbrendű organizmusokból izolált U3 RNS-ek 5' végeinek másodlagos szerkezete. (A) Dictyostelium discoideum U3, (B) Saccharomyces cerevisiae U3b, (C) Schizosaccharomyces pombe U3, (D) S.pombe U3 Porter és mtsai (1988) szerinti másodlagos szerkezet. Az egyes stem-ek mellett feltüntettük a PCFOLD-dal kapott szabadenergia-csökkenés értékeket. A C és D ábrán a zárójelben lévő nukleotidok az S.pombe U3 egy újonnan megszekvenált variánsában lévő változásokat mutatják.

másodlagos szerkezet. Ennél kisebb kompromisszummal is konszenzusra lehet azonban jutni: a magasabbrendű állati U3 RNS-ek 5' vége is beleilleszthető a növényi és alacsonyabbrendű eukariótákból származtatott 2 hajtűhurkos másodlagos szerkezetbe (4.15. ábra). Igaz, hogy ennek következtében e gerinces U3 RNS-ek jelentősen veszítenek stabilitásukból (a humán U3 esetében például ha az 5' végi egy hajtűhurkot átalakítjuk két hajtűhurkos szerkezetté, akkor -21.9 kcal/mol helyett csak -5 kcal/mol lesz erre a régióra a szabadenergia csökkenés), viszont egyrészt éppen a stabilitáscsökkenés, másrészt a szerkezetátalakítás következtében (a konzervált A box hurokba kerülésével) e régió alkalmasabbá válhat más molekulákkal való kölcsönhatásra. Ez jelentheti egyrészt fehérjék hozzákötődését a hajtű-hurokhoz, másrészt más RNS-ekkel való bázispárosodást (amit nyilván elősegíthet a stabilitáscsökkenés, hiszen kisebb stabilitású másodlagos szerkezet könnyebben fel tud bomlani, és egy másik RNS-sel kölcsönhatásba tud lépni).

4.3.4. Az U3 RNS funkciója

Jórészt az U1 és az U2 RNS analógiájára, melyek célszekvenciáikat bázispárosodás révén ismerik föl, a pre-rRNS érésére is számos, az U3 és a pre-rRNS közötti bázispárosodáson alapuló modell született.

Az első ilyen (még teljesen hipotetikus) modell Bachellerie és mtsai (1983) nevéhez fűződik, aki feltételezte, hogy a C box bázispárosodás



4.15. ábra: A humán U3 RNS 5' végének módosított másodlagos szerkezete az energiaértékek feltüntetésével.

révén ismeri föl az ITS 2 szekvenciát, lehetővé téve ezzel a határoló 5.8 S rRNS és rRNS közötti hélix kialakulását.

Egy másik (ugyancsak hipotéziseken alapuló) elmélet szerint (Parker és Steitz, 1987) a B box ismeri föl bázispárosodás révén a riboszomális pre-28S RNS 3' végét, és ilymódon annak érésében játszik szerepet.

Mindkét modellel szemben felmerül az az ellenvetés, hogy a filogenetikai konzerváltság hiánya miatt nem alkalmazható általánosan. A Bachellerie-modell érvényességét Porter és mtsai kérdőjelezik meg alacsonyabbrendűek vizsgálata alapján, a Parker-modell érvényessége pedig már növényekre sem terjed ki (paradicsom 25S RNS génjének 3'

határoló szekvenciája alapján, Kis Mihály személyes közlése).

Legutóbb Stroke és Weiner (1989) illetve Maser és Calvet (1989) egymástól függetlenül pszoralénes keresztkötési kísérletekből azt találták, hogy az U3 RNS a 18S rRNS-től 5' irányban lévő ETS-hez (External Transcribed Spacer=külső átírt szakasz) kötődik. A keresztkötés feltételezések szerint az U3 Box A régiója és az ETS primér processing helye (ez a hely vesz részt az rRNS érési folyamatának első lépésében, melynek során az ETS 5' vége lehasad az rRNS operonról) között jön létre. Emlős ETS szekvenciák összehasonlításából kitűnt (Stroke és Weiner, 1989), hogy e szekvenciák valóban konzerváltak a primér processing hely közelében (4.16. ábra), e konzervált régió azonban nem mutat komplementaritást az U3 A boxával. A szerzők a patkány ETS-ben ettől a helytől csak 80 nukleotiddal arrébb (3' irányban) találtak komplementaritást az A boxszal, amelynek érvényessége azonban más emlősökre már nem terjedt ki.

A processing hely szekvenciájának konzerváltságát ellenőrizendő, az elérhető növényi ETS-ek szekvenciáját részletes komputeres analízisnek vetettem alá. (A szekvenciák forrásai: *Vicia faba* [lóbab] és *Pisum sativum* [borsó]: Kato és mtsai, 1990; *Lycopersicum esculantum*

```
Rat    (787)          CCCGUUG--GAUCGAUGUGGUGCUGCCGC
                       *  **   ************** * *
Mouse  (650)          ACUCUUA--GAUCGAUGUGGUGCUCCGGA
                       *****   *************      *
Human  (415)          GCUCUUCGUGAUCGAUGUGGUGACGUCGU
```

4.16. ábra:Az emlős ETS szekvenciák konzerváltak a riboszomális RNS prekurzor molekula primér processing helyének közelében (Stroke és Weiner /1989/ után).

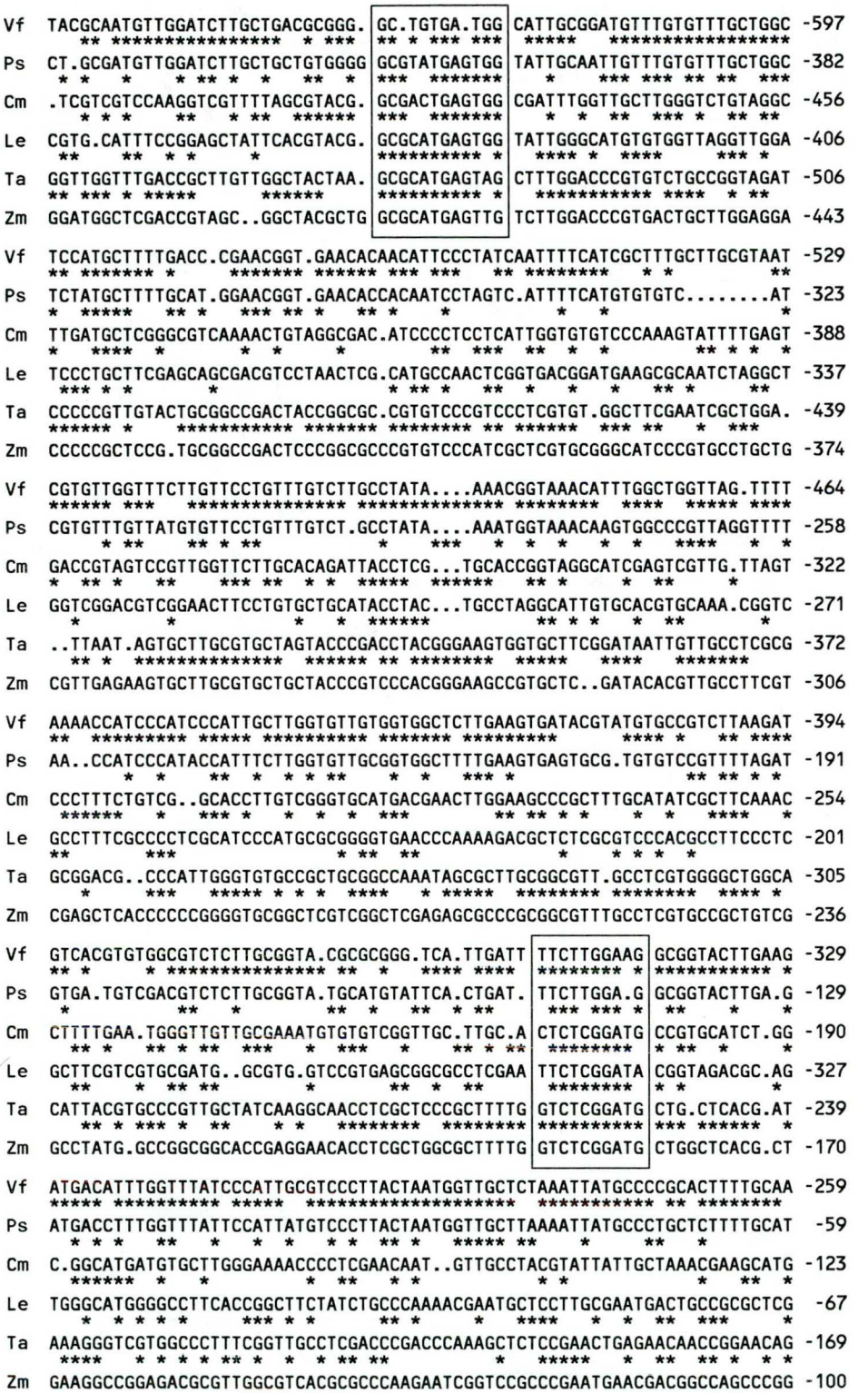
[paradicsom]: Perry és mtsai, 1990; *Triticum aestivum* [búza]: Barker és mtsai (1988); *Zea mays* [kukorica]: McMullen és mtsai (1986); *Cucurbita maxima*: EMBL adatbank.) A páronkénti összehasonlítások alapján (főleg a mátrixos D3HOM adott releváns eredményeket) származtatott többszörös illesztésből kitűnik (4.17. ábra), hogy a növényi ETS szekvenciákban két, kolineárisan elhelyezkedő konzervált box is található, de ezek egyike sem egyezik meg az emlős processing hely konzervált szekvenciájával (és természetszerűleg az U3 RNS szekvenciájával sem komplementer).

Mindezek alapján a pszoralénes keresztkötési kísérletek eredményei inkább a molekulák térbeli közelségére utalnak, semmint kiterjedt bázispárosodásra (Savino és Gerbi, 1990).

Kass és mtsai (1990) immunoprecipitációs és nukleázzal végzett kísérletekből úgy találták, hogy az U3 RNS nélkülözhetetlen az ETS-ben történő primér processing lépéshez, bár meglepő módon az A boxszal komplementer oligonukleotidnak nem volt hatása az ETS hasítására.

Bonyolítják a képet Savino és Gerbi (1990) vizsgálatai, akik Xenopus U3 RNS in vivo hasítási vizsgálataiból arra a következtetésre jutottak, hogy a riboszomális RNS érése Xenopusban két, egymástól független úton történik, de meglepő módon mindkét út esetében az U3 az ITS 1 - 5.8S rRNS hasítási lépésében vesz részt.

Mivel a fenti modellek egymásnak ellentmondóak és filogenetikailag sem konzerváltak, ezért mi egy olyan modellt konstruáltunk, amely az utóbbi filogenetikai feltételnek eleget tesz, és a kísérleti eredmények zömével is összhangba hozható:



4.17. ábra:Növényi ETS szekvenciák többszörös illesztése. (Rövidítések: Vf, Vicia faba; Ps, Pisum sativum; Le, Cm, Cucurbita maxima; Lycopersicum esculentum; Ta, Triticum aestivum) A bekeretezéssel a két legkonzerváltabb régiót jelöltük, a számozások megfelelnek a 18S rRNS-ek 5' végéhez képesti pozícióknak.

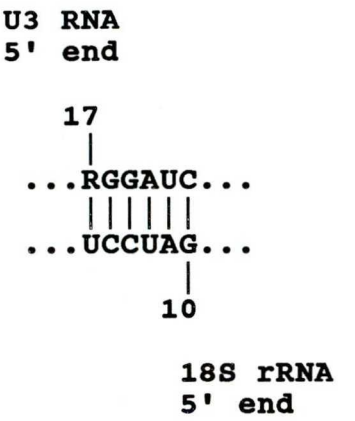
Feltételeztük, hogy ha az U3 bázispárosodás révén vesz részt a riboszomális RNS érési folyamataiban, akkor azt az A box révén teszi. Erre a feltételezésre a következők indítottak:

i/ mint láttuk, az U3 legkonzerváltabb régiója az A box,

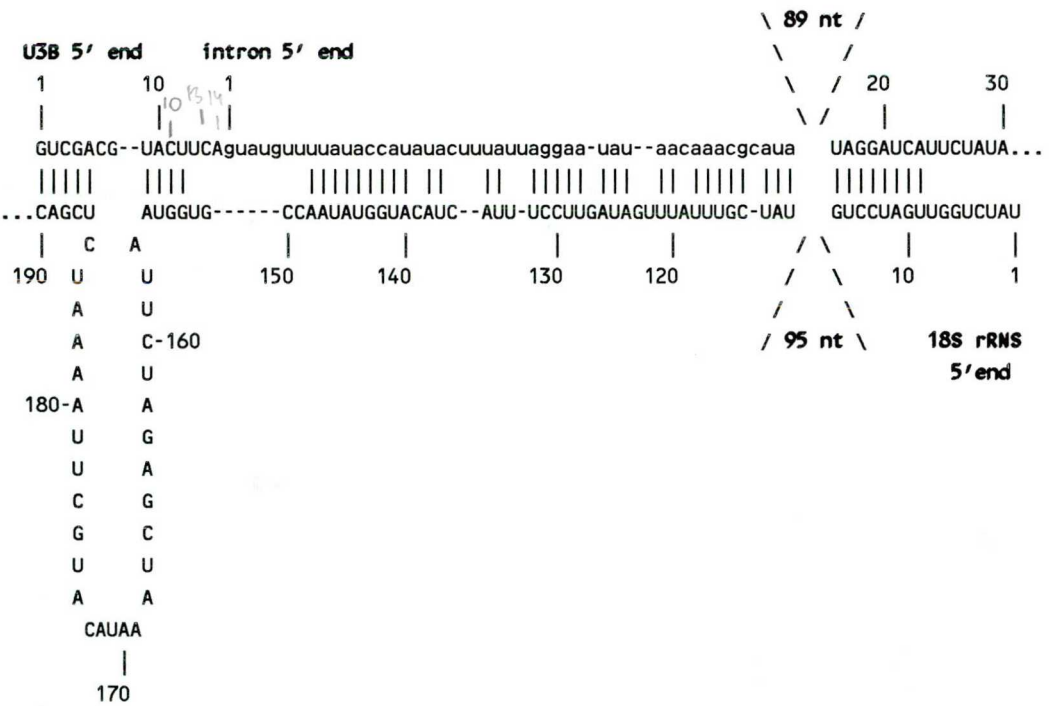
ii/ a molekulának ebben a régiójában van a legtöbb szabadsági foka más molekulákkal való interakcióra (itt tud a legkönnyebben felbomlani a másodlagos szerkezet, ha egyáltalán létezik). Megfigyelhető, hogy az összes többi, más molekulákkal való bázispárosodásban részt vevő U RNS-nek kifejezetten egyszálú régiója vesz részt az ilyenfajta kölcsönhatásban (ebből a szempontból a hurok sem megengedett). Így például az U1 egyesszálú 5' vége kapcsolódik az intron 5' vágási helyéhez, az U2 egyszálú régiója kapcsolódik az intron elágazási pontjához (Keller és Noon, 1985), az U4 egyesszálú 5' vége bázispárosodik az U6-tal (Guthrie és Patterson, 1988), sőt, a legújabb adatok szerint az U2 5' vége is bázispárosodik az U6-tal (Wu és Manley, 1991; Datta és Weiner, 1991) (a modellt kompenzatórikus báziscserék is alátámasztják, Solymosy Ferenc személyes közlése).

Mivel az ETS-ben nem találtunk az A boxhoz hasonló konzerváltságú régiót, ezért magában a riboszomális RNS-ben kerestünk: a 18S rRNS 5' végében találtunk is egy filogenetikailag teljesen konzervált régiót, ami 6 nukleotidon keresztül teljesen komplementer az A box 6 nukleotidjával (4.18. ábra).

További érv ennek a riboszomális RNS régiónak az involváltságára az a komplementaritás, amit az élesztő (*Saccharomyces cerevisiae*) U3B RNS génjében talált intron (Myslinski és mtsai, 1990) és az élesztő 18S rRNS szekvenciája között fedeztünk fel (4.19. ábra).



4.18. ábra: Az U3 RNS 5' vége (Box A régiója) és a 18S rRNS 5' vége közötti bázispárosodási modell. A modell az összes ismert U3 RNS és 18S rRNS szekvenciájára érvényes (a számozás a paradicsom szekvenciáinak felel meg).



4.19. ábra: Az élesztő U3B intronja és az élesztő 18S rRNS 5' vége közötti komplementaritás. Az intron komplementer részét a kisbetűs szekvencia jelöli, a nem komplementer részeket (az intronban 89, a 18S rRNS-ben 95 nukleotidot) nem írtuk ki, csak számmal jelöltük. Látható, hogy komplementaritás - mindkét irányban - az intronon túl is található.

Az intron az U3 szekvenciájába a 13. és a 14. pozíció között ékelődött be. Eredetére vonatkozólag Myslinski és mtsai semmit sem tudtak mondani, csak annyit jegyeztek meg, hogy "mivel az U3 nukleoláris RNS, amely a riboszomális RNS éréséhez szükséges, ezért az intronját nem szerezhette valamilyen abnormális pre-messenger RNS érési folyamat során", eltérően az *S.pombe* U6 RNS intronjától (Tani és Ohshima, 1989) amely Brow és Guthrie (1989) szerint pontosan így keletkezhetett.

Feltételezéseink szerint az élesztő U3 intronját a riboszomális RNS processingje során szerezte, mégpedig úgy, hogy miután az U3 RNS és a 18S RNS bázispárosodással összekapcsolódtak, egy polimeráz átírta az U3 RNS szekvenciájába. Ilyen polimeráz található a legtöbb *S.cerevisiae* törzsben jelenlévő kettősszálú RNS retrovírusokban, melyek e polimerázuk segítségével önállóan képesek szaporodni az élesztő sejtekben (Wickner, 1989).

5. ÖSSZEFOGLALÁS

Munkánk fő célkitűzése a kis RNS-ek (ezen belül a sejtmagi kis RNS-ek és a viroidok) primér szekvenciájára és másodlagos szerkezetére vonatkozó törvényszerűségek feltárása, és az azokból levonható funkcionális következtetések származtatása volt. Ezek a következők:

(i) A viroidok primér szekvenciájának strukturális vizsgálatára kifejlesztettünk egy számítógépes programot, melynek a segítségével kimutattuk, hogy az összes viroidok szekvenciája periodikus tulajdonságokat mutat:

a/ a PSTV-ben és a CCCV-ben és a velük rokon viroidokban egy lazán definiált, 12 nukleotidos periódussal és azonos konszenzus szekvenciával jellemezhető periodicitás található,

b/ az ASBV-t egy 80 nukleotidos periodicitás jellemzi, amely a 247 nt hosszú szekvenciát három homológ részre osztja,

c/ az ASSV-ben egy 60 nukleotidos periodicitás található.

A viroidok szekvenciáinak ezen egyedülálló periodikus sajátosságai arra engedtek következtetni, hogy e periodicitás a viroidok fehérjekötő kapacitásának növekedését segíti elő, amelynek aztán fontos szerepe lehet a viroidok transzkripciójában és/vagy patogén hatásának kifejtésében.

(ii) A rendelkezésünkre álló nyolc paradicsom U1 gén szekvenciáiból evolúciós fákat származtattunk. Megállapítottuk, hogy a gént kódoló és az azokat határoló szekvenciákból különböző módszerekkel különböző törzsfák származtathatók, amit két tényezőnek tulajdonítottunk,

1/ az eltérő szelekciós nyomásnak, amelynek a különböző régiók ki voltak téve,

2/ a tranzícióknak és a transzverzióknak az egyes régiókban előforduló eltérő gyakoriságának.

Megállapítottuk, hogy bár az emlős U1 pszeudogének esetében is hasonló eltérés mutatkozik a tranzíciók és transzverziók között, mint a paradicsom U1 gének esetében, azok másodlagos szerkezeti eloszlása jelentősen különbözik a két géncsoportban, ezzel is alátámasztva azt a feltevést, hogy a paradicsom U1 gének valódi, expresszálódó gének (noha kísérletileg csak az U1.1 variáns expresszióját mutatták ki).

Meghatároztuk a paradicsom U3 RNS energetikailag optimális másodlagos szerkezetét. Megállapítottuk, hogy az lényegében megegyezik az alacsonyabbrendű organizmusokból izolált U3 RNS másodlagos szerkezetével. Noha a magasabbrendű állatok U3 RNS-ének 5' vége másodlagos szerkezetét tekintve jelentősen különbözik a növényekétől és az alacsonyabbrendű állatokétól, az ebben a régióban talált nagyfokú primér szekvenciahomológia mégis arra indított bennünket, hogy a magasabbrendű állatok U3 RNS-ének másodlagos szerkezetét is beleillesszük a többi U3 alapján konstruált konszenzusba.

Az ismert növényi ETS szekvenciák többszörös illesztésével két konzervált régiót találtunk, amelyek egyike sem egyezett meg azonban az állati ETS-ek szekvenciájában talált konzervált processing hely szekvenciájával, így azokat a modelleket, amelyek az U3-nak a preriboszomális RNS érésében játszott szerepére az U3 és az ETS szekvenciája közötti bázispárosodást feltételezték, elvetettük. Megállapítottuk viszont, hogy az U3 legkonzerváltabb régiója és a 18S

rRNS 5' vége között evolúciósan teljesen konzervált módon 6 nukleotid hosszúságú komplementaritás található, amelynek a pre-rRNS érésében játszott szerepét az élesztő U3 intronja és a 18S rRNS szekvenciája között talált komplementaritás is alátámasztja.

Ezzel tehát sikerült az U3 RNS másodlagos szerkezetére és a pre-rRNS érésében játszott szerepére egy egységes modellt konstruálni.

6. IRODALOMJEGYZÉK

Abel, S., Kiss, T., and Solymosy, F. (1989): Molecular analysis of eight U1 RNA gene candidates from tomato that could potentially be transcribed into U1 RNA sequence variants differing from each other in similar regions of secondary structure.
Nucleic Acids Res., **17**:6319-6337.

Bachellerie, J.-P., Michot, B. and Raynal, F. (1983): Recognition signals for mouse pre-rRNA processing: A potential role for U3 nucleolar RNA.
Mol. Biol. Rep., **9**:79-86.

Barker, R.F., Harberd, N.P., Jarvis, M.G. and Flavell, R.B. (1988): Structure and evolution of the intergenic region in a ribosomal DNA repeat unit of wheat.
J. Mol. Biol., **201**:1-17.

Black, D. L., Chabot, B., and Steitz, J.A. (1985): U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing.
Cell, **42**:737-750.

Black, D.L. and Steitz, J.A. (1986): Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein.
Cell, **46**:697-704.

Borer, P.N., Dengler, B., Tinoco, I. and Uhlenbeck, O.C. (1974): Stability of ribonucleic acid double-stranded helices.
J. Mol. Biol., **86**:843-853.

Branch, A.D. and Robertson, H.D. (1984): A replication cycle for viroids and other small infectious RNAs.

Science, **223**:450-455.

Branch, A.D., Robertson, H.D. and Dickson, E. (1981): Longer-than-unit length viroid minus strands are present in RNA from infected plants.

Proc. Natl. Acad. Sci. USA, **78**:6381-6385.

Branlant, C., Krol, A., Ebel, J.-P., Lazar, E., Haendler, B., and Jacob, M. (1982): U2 RNA shares a structural domain with U1, U4, and U5 RNAs.

EMBO J., **1**:1259-1265.

Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, M. and Lührmann, R. (1984): Evidence for existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing.

EMBO J., **3**:1357-1363.

Brow, D.A. and Guthrie, C. (1989):RNA-processing - splicing a spliceosomal RNA.

Nature, **337**:14.

Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C. (1982):Mitochondrial DNA sequences of primates: tempo an mode of evolution.

J. Mol. Evol., **18**:225-239.

Brunel, C., Sri-Widada, J., Lelay, M.N., Jeanteur, P. and Liautard, J.P. (1981): Purification and characterization of a simple ribonucleoprotein particle containing small nucleoplasmic RNAs (snRNP) as a subset of RNP containing heterogenous nuclear RNP (hnRNP) from HeLa cells.

Nucleic Acids Res., **9**:815-830.

Calvet, J.P. and Pederson, T. (1981): Base-pairing interactions between small nuclear RNAs and nuclear RNA precursors as revealed by psoralen cross-linking in vivo.

Cell, **26**:363-370.

Cabot, E.L. and Beckenbach, A.T. (1989): Simultaneous editing of multiple nucleic acid and protein sequences with ESEE.

CABIOS, **5**:233-234.

Chabot, B., Black, D.L., LeMaster, D.M. and Steitz, J.A. (1985): The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribosomal ribonucleoprotein.

Science, **230**:1344-1349.

Corpet, F. (1988): Multiple sequence alignment with hierarchical clustering.

Nucleic Acids Res., **16**:10881-10890.

Crick, F., (1979): Split genes and RNA splicing.

Science, **204**:264-271.

Crouch, R.J., Kanaya, S. and Earl, P.L. (1983): A model for the involvement of the small nuclear RNA (U3) in processing eukaryotic ribosomal RNA.

Mol. Biol. Rep., **9**:75-78.

Delisi, C. and Crothers, D.M. (1971): Prediction of RNA secondary structure.

Proc. Natl. Acad. Sci. USA, **68**:2682-2685.

Dahlberg, J. E. and Lund, E., (1988): The genes and transcription of the major small nuclear RNAs, in *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles,* Birnstiel, M. L., Ed., Springer Verlag, Berlin, p. 38-70.

Datta, B. and Weiner, A.M. (1991): Genetic evidence for base pairing between U2 and U6 snRNA in mammalian mRNA splicing. *Nature,* **352**:821-824.

Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986): Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA,* **83**:9373-9377.

Fristensky, B., Lis, J. and Wu, R. (1982): Portable microcomputer software for nucleotide-sequence analysis. *Nucleic Acids Res.,* **10**:6451-6463.

Gralla, J. and Crothers, D.M. (1973a): Free energy of imperfect nucleic acid helices. II. Small hairpin loops. *J. Mol. Biol.,* **73**:497-511.

Gralla, J. and Crothers, D.M. (1973b): Free energy of imperfect nucleic acid helices. III. Small internal loops resulting from mismatches. *J. Mol. Biol.,* **78**:301-319.

Guthrie, C. and Patterson, B. (1988): Spliceosomal snRNAs. *Annu. Rev. Genet.*, **22**:387-419.

Hashimoto, C. and Steitz, J.A. (1984): U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res.,* **12**:3283-3293.

Hashimoto, J. and Koganezawa, H. (1987):Nucleotide-sequence and secondary structure of apple scar skin viroid. *Nucleic Acids Res.,* **15**:7045-7052.

Herr, W. and Clarke, J. (1986): The SV40 enhancer is composed of multiple functional elements that can compensate for one

another.

Cell, **45**:461-470.

Hodnett, J.L. and Busch, H. (1968): Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei.

J. Biol. Chem., **243**:6334-6342.

Hughes, J.M.X., Konings, D.A.M., and Cesareni, G. (1987): The yeast homologue of U3 snRNA.

EMBO J., **6**:2145-2155.

Jeppesen, C., Stebbins-Boaz, B. and Gerbi, S.A. (1988): Nucleotide sequence determination and secondary structure of Xenopus U3 snRNA.

Nucleic Acids Res., **16**:2128-2148.

Kass, S., Tyc, K., Steitz, J. A., and Sollner-Webb, B. (1990): The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing.

Cell, **60**:897-908.

Kato, A., Nakajima, T. Yamashita, J. Yakura, K. and Tanifuji, S. (1990): The structure of the large spacer region of the rDNA in *Vicia faba* and *Pisum sativum*.

Plant Mol. Biol., **14**:983-993.

Keese, P. and Symons, R.H. (1985): Domains in viroids: Evidence of intermolecular RNA rearrangements and their contribution to viroid evolution.

Proc. Natl. Acad. Sci. USA, **82**:4582-4586.

Keller, E.B. and Noon, W.A. (1985): Intron splicing: a conserved internal signal in introns of *Drosophila* pre-mRNAs.

Nucleic Acids Res., **13**, 4971-4981.

Kierzek, R., Caruthers, M.H., Longfellow, C.E., Swinton, D., Turner, D.H. and Freier, S.M. (1986): Polymer-supported RNA-synthesis and its application to test the nearest-neighbor model for duplex stability.

Biochemistry, **25**:7840-7846.

Kiss, T. (1987): Kandidátusi értekezés.

Kiss, T. and Solymosy, F. (1990): Molecular analysis of a U3 RNA gene locus in tomato: transcription signals, the coding region, expression in transgenic tobacco plants and tandemly repeated pseudogenes.

Nucleic Acids Res., **18**:1941-1949.

Kiss, T. and Solymosy, F. (1982): Sequence homologies between a viroid and a small nuclear RNA (snRNA) species of mammalian origin.

FEBS Lett., **144**:318-320.

Kiss, T., Abel, S., and Solymosy, F. (1989): A plant pseudogene for U1 RNA.

Plant Mol. Biol., **12**:709-711.

Kiss, T., Antal, M., Hegyi, H. and Solymosy, F. (1988a): Plant small nuclear RNAs. IV. The structure of U1 RNA from *Chlorella saccharophila*: a phylogenetic support, in terms of RNA structure, for the probable interaction between U1 and U2 RNPs during the splicing of pre-mRNA.

Nucleic Acids Res., **16**:2734.

Kiss, T., Jakab, G., Antal, M., Pálfi, Z., Hegyi, H., Kis, M., and Solymosy, F. (1988b): Plant small nuclear RNAs. V. U4 RNA is present in broad bean plants in the form of sequence variants and is

base-paired with U6 RNA.

Nucleic Acids Res., **16**:5407-5426.

Kiss, T., Pósfai, J., and Solymosy, F. (1983): Sequence homology between potato spindle tuber viroid and U3B snRNA.

FEBS Lett., **163**:217-220.

Kiss, T., Tóth, M., and Solymosy, F. (1985): Plant small nuclear RNAs. Nucleolar U3 snRNA is present in plants: partial characterization.

Eur. J. Biochem., **152**:259-266.

Kluge, A.G. and Farris, J.S. (1969): Quantitative phyletics and the evolution of anurans.

Systematic Zoology, **18**:1-32.

Koltunow, A.M. and Rezaian, M.A. (1989):A scheme for viroid classification.

Intervirology, **30**:194-201.

Korn, L.J., Qeen, C.L. and Wegman, M.N. (1977): Computer-analysis of nucleic-acid regulatory sequences.

Proc. Natl. Acad. Sci. USA, **74**:4401-4405.

Krol, A., Branlant, C., Lazar, E., Gallinaro, H., and Jacob M. (1981): Primary and secondary structure of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs.

Nucleic Acids Res., **9**:2699-2716.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980): Are snRNPs involved in splicing?

Nature, **283**:220-224.

Lührmann, R. (1990): Functions of U-snRNPs.

Mol. Biol. Rep., **14**:183-192.

SZTE Egyetemi Könyvtár SZEGED

Lührmann, R. (1988): snRNP proteins, in *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles,* Birnstiel, M. L., Ed., Springer Verlag, Berlin, pp. 71-99.

Maser, R.L. and Calvet, J.P. (1989): U3 small nuclear RNA can be psoralen-cross-linked in vivo to the 5' external trascribed spacer of pre-ribosomal-RNA.
Proc. Natl. Acad. Sci. USA, **86**:6523-6527.

McMullen, M.D., Hunter, B., Phillips, R.L. and Rubenstein, I. (1986):
Nucleic Acids Res., **14**:4953-4968.

Mohamed, N.A. and Thomas, W. (1980): Viroid-like properties of an RNA species associated with the sunblotch disease of avocados.
J. Gen. Virol., **46**:157-167.

Muramatsu, M. and Busch, H. (1964): Studies on nucleolar RNA of the Walker 256 carcinosarcoma and the liver of the rat.
Cancer Res., **24**:1028-1034.

Myslinski, E., Ségault, V., and Branlant, C. (1990): An intron in the genes for U3 small nucleolar RNAs of the yeast *Saccharomyces cerevisiae.*
Science, **247**:1213-1216.

Needleman, S.B. and Wunch, C.D. (1970): A general method applicable to the search for similarities in the amino acid sequence of two proteins.
J. Mol. Biol., **48**:443-453.

Neefs, J.-M., Van de Peer, Y., Hendriks, L. and DeWachter, R. (1990): Compilation of small ribosomal subunit RNA sequences.
Nucleic Acids Res., **18**:2237-2318.

Nussinov, R. and Jacobson, A.B. (1980): Fast algorithm for predicting the

secondary structure of single-stranded RNA.

Proc. Natl. Acad. Sci. USA, **77**:6309-6313.

Parker, K. A. and Steitz, J. A. (1987): Structural analyses of the human U3 ribonucleoprotein particle reveal a conserved sequence available for base pairing with pre-rRNA.

Mol. Cell. Biol., **7**:2899-2913.

Perry, K.L. and Palukaitis, P. (1990): Transcription of tomato ribosomal DNA and the organization of the intergenic spacer.

Mol. Gen. Genet., **221**:102-112.

Pipas, J. and MacMahon, J. (1975): Method for predicting RNA secondary structure.

Proc. Natl. Acad. Sci. USA, **72**:2017-2021.

Pleij, C.W.A., Rietveld, K. and Bosch, L. (1985): A new principle of RNA folding based on pseudoknotting.

Nucleic Acids Res., **13**:1717-1731.

Porter, G.L., Brennwald, P.J., Holm, K.A., and Wise, J.A. (1988): The sequence of U3 from *Schizosaccharo-myces pombe* suggests structural divergence of this snRNA between metazoans and unicellular eukaryotes.

Nucleic Acids Res., **16**:10131-10152.

Prestayko, A.W. and Busch, H. (1968): Low molecular weight RNA of the chromatin fraction from Novikoff hepatoma and rat liver nuclei.

Biochim. Biophys. Acta, **169**:327-337.

Prestayko, A.W., Tonato, M. and Busch, H. (1970): Low molecular weight RNA associator with 28S nucleolar RNA.

J. Mol. Biol., **47**:505-515.

Riesner, D. and Gross, H.J. (1985): Viroids.

Ann. Rev. Biochem., **54**:531-564.

Riesner, D., Henco, K. Rokohl, U., Klotz, G., Kleinschmidt, A.K. et al. (1979): Structure and structure formation of viroids.

J. Mol. Biol., **133**:85-115.

Rinke, J., Appel, B., Blöcker, M., Frank, R. and Lührmann, R. (1984): The 5'-terminal sequence of U1 RNA complementary to a consensus 5' splice site of hnRNP is single-stranded in intact U1 snRNP particles.

Nucleic Acids Res., **12**:4111-4126.

Rinke, J., Appel, B., Digweed, M., and Lührmann, R. (1985): Localization of a base-paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen cross-linking.

J. Mol. Biol., **185**:721-731.

Salser, W. (1977): Globin mRNA sequences: Analysis of base-pairing and evolutionary implications.

Cold Spring Harbor Symp. Quant. Biol., **XLII**:985-1004.

Savino, R. and Gerbi, S.A. (1990): *In vivo* disruption of Xenopus U3 snRNA affects ribosomal RNA processing.

EMBO J., **9**:2299-2308.

Schneider, T.D., Stormo, G.D., Gold, L. and Ehrenfeucht, A. (1986): Information content of binding sites on nucleotide sequences.

J. Mol. Biol., **188**:415-431.

Schnölzer, M., Haas, B., Ramm, K., Hoffmann, H. and Sanger, H.L. (1985): Correlation between structure and pathogenicity of potato spindle tuber viroid (PSTV).

EMBO J., **4**:2181-2190.

Sekeris, C.E. and Niessing, J. (1975): Evidence for the existenvce of a structural RNA component in the nuclear ribonucleoprotein particles containing heterogenous RNA.
Biochem.Biophys.Res.Comm., **62**:642-648.

Shannon, C.E. and Weaver, W. (1949): The mathematical theory of communication. University of Illinois Press, Urbana.

Sogo, J.M., Koller, T. and Diener, T.O. (1973): Potato spindle tuber viroid. X. Visuallization and size determination by electron microscopy.
Virology, **55**:70-80.

Sokal, R.R. and Sneath, H.A. (1963): Principles of numerical taxonomy. W.H. Freeman, San Francisco.

Solymosy, F. and Kiss, T. (1985): Viroids and snRNAs, in *Subviral Pathogens of Plants and Animals: Viroids and Prions*, Maramorosch, K. and McKelvey, J. J., Eds., Academic Press, New York, pp. 183.

Solymosy, F. (1990): Are UsnRNA sequence variants involved in developmentally controlled, tissue-specific and/or alternative splicing of pre-mRNA at the level of RNA-protein interaction? Some hints from studies of plant systems.
Mol. Biol. Rep., **14**:145.

Solymosy, F. (1990): Similarities and dissimilarities between plant UsnRNAs and their equivalents in other eukaryotes with respect to structure, complexing with proteins, gene organization and function.
Acta Biochim. Biophys. Hung., **25**:67-85.

Steitz, J.A., Black, D.L., Gerke, V., Parker, K.A., Krämer, A., Frendewey, D., and Keller, W., Functions of the abundant U-snRNPs, in *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles,* Birnstiel, M. L., Ed., Springer Verlag, Berlin, 1988, 115.

Stroke, I.L. and Weiner, A.M. (1989): The 5' end of U3 snRNA can be crosslinked *in vivo* to the external transcribed spacer of rat ribosomal RNA precursors.
J. Mol. Biol., **210**:497-512.

Studnicka, G.M., Rahn, G.M., Cummings, I.W. and Salser, W.A. (1978): Computer method for predicting the secondary structure of single-stranded RNA.
Nucleic Acids Res., **5**:3365-3387.

Tague, B.W. and Gerbi, S.A. (1984): Possible function of U3 RNA.
J. Mol. Evol., **20**:362-367.

Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973): Improved estimation of secondary structure in ribonucleic acids.
Nature New Biol. **246**:40-41.

Turner, D.H., Sugimoto, N., Jaeger, J.A., Longfellow, C.E., Freier, S.M. and Kierzek, R. (1987): Improved parameters for prediction of RNA structure.
Cold Spring Harbor Symp. Quant. Biol., **LII**:123-133.

Tyc, K. and Steitz, J.A. (1989): U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus.
EMBO J., **8**:3113-3119.

van Santen, V.L. and Spritz, R.A. (1987): Nucleotide sequence of a bean

(*Phaseolus vulgaris*) U1 small nuclear RNA gene: Implications for plant pre-mRNA splicing.
Proc. Natl. Acad. Sci. U.S.A., **84**:9094-9098.

van Santen, V.L., Swain, W., and Spritz, R.A. (1988): Nucleotide sequences of two soybean U1 snRNA genes.
Nucleic Acids Res., **16**:4176.

Vankan, P. and Filipowicz, W. (1989): A U-snRNA gene-specific upstream element and a -30 'TATA box' are required for transcription of the U2 snRNA gene of *Arabidopsis thaliana.*
EMBO J., **8**:3875-3882.

Vankan, P. and Filipowicz, W. (1988): Structure of U2 snRNA genes of *Arabidopsis thaliana* and their expression in electroporated plant protoplasts, *EMBO J.*, **7**:791-799.

Vankan, P., Edoh, D., and Filipowicz, W. (1988): Structure and expression of the U5 snRNA gene of *Arabidopsis thaliana*. Conserved upstream sequence elements in plant U-RNA genes.
Nucleic Acids Res., **16**:10425-10440.

Vincent, A. (1986): TFIIIA and homologous genes. The 'finger' proteins.
Nucleic Acids Res., **14**:4385-4391.

Weinberg, R. A. and Penman, S.: Small molecular weight monodisperse nuclear RNA.
J. Mol. Biol., **38**:289-304.

Wickner, R.B. (1989): Yeast virology.
Faseb J., **3**:2257-2265.

Wilbur, W.J. and Lipman, D.J. (1983): Rapid similarity searches of nucleic acid and protein data banks
Proc. Natl. Acad. Sci. USA, **80**:726-730.

Williams, A.L. and Tinoco, I. (1986): A dynamic programming algorithm for finding alternative RNA secondary structures.
Nucleic Acids Res., **14**:299-315.

Wise, J.A. and Weiner, A.M. (1980): Dictyostelium small nuclear RNA D2 is homologous to rat nucleolar RNA U3 and is encoded by a dispersed multigene family.
Cell, **22**:109-118.

Wolff, P., Glitz, R., Schumacher, D. and Riesner, D. (1985): Complexes of viroids with histones and other proteins.
Nucleic Acids Res., **13**:355-366.

Wu, J. and Manley, J.L. (1991): Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA.
Nature, **352**:818-821.

Yamamoto, K. and Yoshikura, H. (1985): Computer program for prediction of the optimal and suboptimal secondary structures of long RNA molecules.
CABIOS, **1**:89-94.

Zhuang, Y. and Weiner, A.M. (1986): A compensatory base change in U1 snRNA suppresses a 5' splice site mutation.
Cell, **46**:827-835.

Zhuang, Y., Goldstein, A.M., and Weiner, A.M. (1989): UACUAAC is the preferred branch site for mammalian mRNA splicing.
Proc. Natl. Acad. Sci. U.S.A., **86**:2752-2756.

Zuker, M. and Stiegler, P. (1981): Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information.
Nucleic Acids Res., **9**:133-148.

7. KÖSZÖNETNYILVÁNÍTÁS

Köszönettel tartozom dr. Solymosy Ferencnek, hogy lehetővé tette a vezetése alatt álló csoportban doktori dolgozatom elkészítését. Köszönöm a munkám során nyújtott értékes elméleti segítségét.

Köszönetet mondok munkatársamnak, Juhász Andrásnak értékes szakmai tanácsaiért és segítőkészségéért.