

RNS tájképek

Dr. Kun Ádám
tudományos főmunkatárs

ELTE Növényrendszertani, Ökológia és Elméleti
Biológia Tanszék

2013.

A rátermettség tájképek, azaz a lehetséges genotípusokat fenotípushoz, s így rátermettséghez rendelő függvények, szerkezete fontos az evolúcióbiológia szempontjából. Fontossága ellenére igen keveset tudunk róluk, mert adatok gyűjteni egy ilyen tájképről igen nehézkes. Gondoljunk csak bele, hogy maga a rátermettség mérése sem egyszerű, s a teljes genom tér iszonyúan nagy (ez a tulajdonsága teszi lehetővé a nyíl végű evolúciót). Teljes élőlények esetében a rátermettségtájképek pár genotípusra ismerhetők csak meg, mikroorganizmusok esetében talán egy kicsit többen. Egyelőre a molekulák szintjén van lehetőségünk, hogy nagyobb számú mutánst vizsgáljunk, s a rátermettséget egy azzal jól korreláló és mérhető tulajdonsággal arányosítsuk. Egy enzim esetében ez például az enzimaktivitás.

Az élet keletkezése szempontjából, feltételezve az RNS világot, a ribozimok enzimaktivitása az ami fontos komponense minden RNA alapú szervezetnek (vagy proto-szervezetnek). Az enzimaktivitás szempontjából a leginkább meghatározó maga a szerkezet, hiszen az aktív centrumot magába foglaló "zsebnek" ki kell alakulnia. Természetesen az aktív centrumban egyes csoportok pontos elhelyezkedése, s így a pontos összetétel is fontos, de kísérletes eredmények alapján is jó közelítés, hogy amíg egy mutáció a szerkezet nem változtatja, addig az enzimaktivitásra is csekély a hatása.

A biopolimerek kutatásának egyik központi kérdése, hogy egyszerű, de pontos módszerekkel lehessen megállapítani, hogy egy-egy szekvencia milyen háromdimenziós szerkezetbe gombolyodik fel. Az RNS molekulák egyik előnye, hogy a másodlagos struktúra szintjén a struktúra jóslása (összefoglaló Higgs 2000-ben) elfogadható pontossággal megoldható.

Az RNA másodlagos szerkezet számolása *

A másodlagos struktúra az RNS lánc tekeredése közben a molekulán belül egymáshoz közel kerülő, megfelelő bázisok részvételével kialakuló, hidrogénhidakkal rögzített bázispárok listája. Így egy RNS szekvencia ($R=r_1, r_2, r_3, \dots, r_N$; ahol $r_i \in \{A,C,G,T\}$) másodlagos struktúrája (S) az $[i, j]$ bázispárok rendezett halmaza, ahol minden bázispárra, $1 \leq i < j \leq N$ mellett igaz, hogy

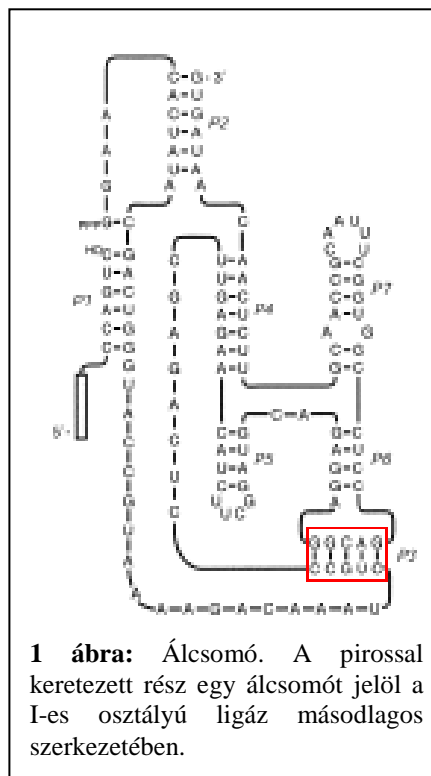
$$j - i > 3 \tag{1}$$

* Ez az alfejezet a teljesség miatt van itt, nem várom el, hogy részleteiben tudjátok.

, és bármely két $[i, j]$ és $[k, l]$ bázispárra igaz, hogy

$$\begin{aligned} i = k &\Leftrightarrow j = l \\ k < j &\Rightarrow i < k < l < j \\ k > j &\Rightarrow i < j < k < l \end{aligned} \quad (2)$$

Az első feltétel alapján egy nukleotid csak egy bázispár kialakításában vehet részt. A második két feltétel biztosítja, hogy a struktúrában nem lehetnek csomók és álcsonok (1. ábra), azaz a másodlagos struktúra egy síkbeli gráfként megjeleníthető. Jóllehet az álcsonok fontosak bizonyos természetes RNS molekulákban is (például a Hepatitis Delta Vírusban), de a harmadlagos struktúra részének tekinthetjük őket. A csomómentes struktúra feltételezése elengedhetetlen ahhoz, hogy dinamikus programozási algoritmussal lehessen a másodlagos struktúrát meghatározni, bár próbálkozások vannak az álcsonok figyelembevételére is (Lyngsø és Pedersen 2000, Isambert és Siggia 2000)



1 ábra: Álcson. A pirossal keretezett rész egy álcsonot jelöl a I-es osztályú ligáz másodlagos szerkezetében.

Az energiamodell

A másodlagos struktúra meghatározásának alapja, hogy a struktúrát egyértelműen hurkokra osztjuk (2. ábra).

Ehhez definiálnunk kell magukat a hurkokat. k -t elérhetőnek nevezzük $[i, j] \in S$ -ből ($i < k < j$), ha nincs olyan $[i', j'] \in S$, amire igaz, hogy $i < i' < k < j' < j$. Ekkor az $[i, j]$ -vel zárt k hurok definíció szerint az $[i, j]$ -ből elérhető $(k-1)$ bázispárból és k' egyszálú bázisból áll. A struktúra így egyértelműen felosztható ilyen k -hurkokra ($L_0, L_1, L_2 \dots L_m$, ahol m a bázispárok száma).

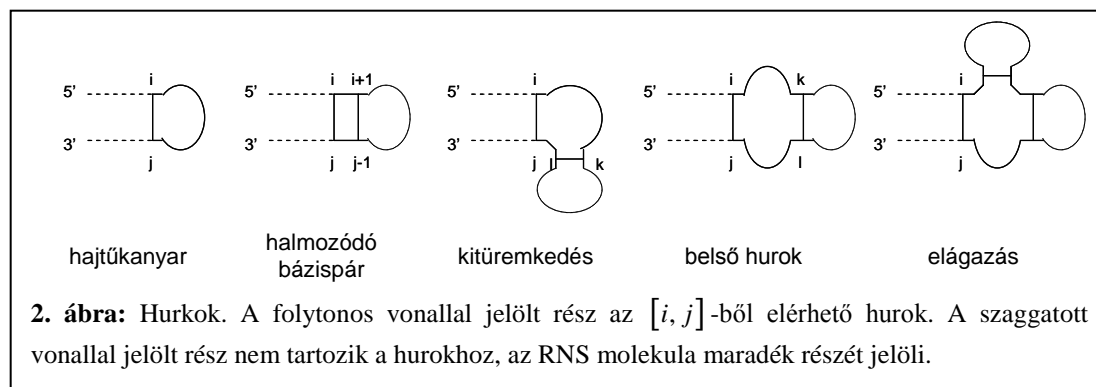
$$R = L_0 \cup \left(\bigcup_{[i, j] \in S} L([i, j]) \right) \quad (3)$$

, ahol

L_0 a $k=0$ hurok, amelybe azok az egyszálú és párt alkotó nukleotidok tartoznak, amelyek egyetlen bázispárból sem elérhetőek (a szekvencia elején vagy végén vannak).

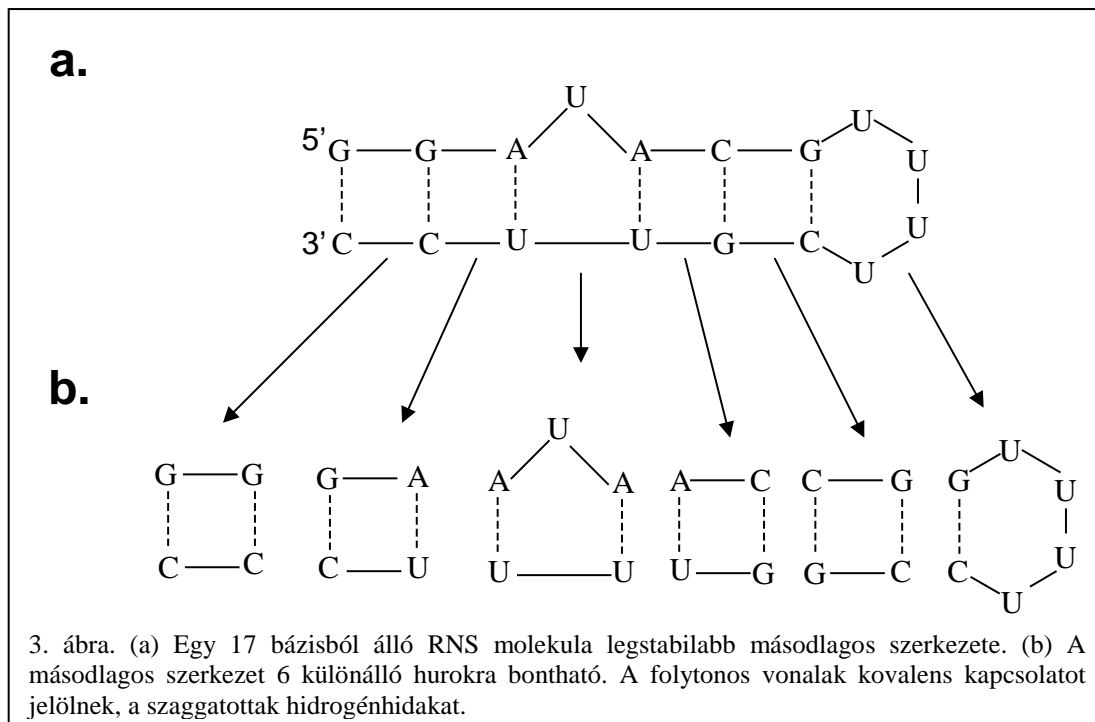
$L([i, j])$ az $[i, j]$ bázispárból elérhető hurok.

Egy RNS molekula másodlagos struktúrájának energiájával kapcsolatban feltesszük, hogy az az egyes hurkok energia hozzájárulásának összege.



$$E(S) = \sum_{i=0}^m e(L_i) \quad (4)$$

Az egyes hurkok energiájárulékát kísérletesen meghatározták, vagy a kísérletes adatokból interpolálták vagy extrapolálták. A legegyszerűbb a halmozódó bázispár energiájáruléka, ami csak a négy résztvevő bázistól függ. A kitüremkedés energiájáruléka alapvetően az $[i, j]$ és a $[k, l]$ bázispártól, valamint a kitüremkedés méretétől függ. A hajtúhurkok energiája a záró bázispártól és a hozzájuk legközelebb levő bázisoktól, a hurok méretétől, és egyes esetekben a hurok konkrét szekvenciájától függ. A belső hurok és az elágazások energiájáruléka ennél is összetettebb, alapvetően azonban főleg a mérettől és a záró bázispártól függ az energia. Az energiamodell részletes leírása Mathews, és *mtsi*. 1999 közleményében található. Az eljárás példázására egy kis molekula energiájának számolását mutatom be. Az 5'-GGAUACGUUUUCGUUGG-3' szekvenciájú RNS molekula legstabilabb másodlagos szerkezete az 3.a ábrán látható. A csomómentes másodlagos szerkezetet általában az úgynevezett zárójelezett reprezentációval írjuk le, amelyben az 5'-végtől kezdve egy pont (.) van a bázis helyén, ha az bázispár kialakításában nem vesz részt, egy nyitó zárójel ((), ha a bázis egy bázispár 5'-véghez közelebbi tagja és egy záró zárójel ()), ha a bázispár 3'-véghez közelebbi tagja. A fenti kis molekula másodlagos



szerkezetét így a $((((\dots))))$ jelsorozattal egyértelműen reprezentálhatjuk.

Az energiaszámításhoz az 3. ábrán bemutatott másodlagos szerkezetet 6 hurokra bontjuk. Ebből négy halmozódó bázispár, egy kitüremkedés és egy hajtúkanyar. A négy halmozódó bázispár energiajáruléka a résztvevő bázisoktól függ és értékük rendre -3,4 kcal, -2,2 kcal, -2,1 kcal és -3,3 kcal. A kitüremkedés energiajáruléka egyrészt a kitüremkedés méretéből ered (1 méretű, energiajáruléka +3,8 kcal), másrészt a két záró bázispár energiajárulékáiból (úgy kell venni, mintha halmozódó bázispárok lennének, így energiajárulékaik -1,1 kcal). A hajtúkanyar energiajáruléka 3 részből tevődik össze: méretéből (4) +5,6 kcal, a lezáró bázispár ([G,C]) és a legközelebbi bázisok (U,U) energiajárulékáiból -0,07 kcal és végül az UUUU kanyar nem tartozik a különösen stabil kanyarak közé, így ebből +0 kcal energia származik. A struktúrát egy [G,C] bázispár zárja le, ami nem von magával energiabüntetést (az [A,U] párral végződők viszont igen). Az egyes energiajárulékok összege, és így a másodlagos struktúra energiája 3,4 kcal/mol.

A szerkezetjósítás a legkisebb energiájú szerkezet megkeresését jelenti. Az energiamodell additív alakjából következően a minimális energia probléma dinamikus programozási algoritmussal megoldható. Az algoritmus hasonló a szekvencia illesztő programok algoritmusához, amelyre Waterman mutatott rá és alkalmazott (Waterman 1978, Waterman és Smith 1978). Az első dinamikus programozási megoldást

Nussinov dolgozta ki (Nussinov és Jacobson 1980, Nussinov, *és mtsi.* 1978), eredetileg olyan struktúra keresésére, amelyben maximális számú bázispár van. A ma is alkalmazott energiamodellel az első minimum energia probléma megoldására használható algoritmust Zuker és Stiegler (Zuker és Sankoff 1984, Zuker és Stiegler 1981) fejlesztette ki.

A módszer számításgénye a szekvencia N hosszával $O(n^3)$ szerint, a memória igénye pedig $O(n^2)$ nő.

Az algoritmus egy ingyen hozzáférhető implementációja a **Vienna RNA Package** (Hofacker, *és mtsi.* 1994), ami elérhető a

<http://www.tbi.univie.ac.at/~ivo/RNA/>

címen. A szakdolgozatban másodlagos struktúra becslésre ezt a programcsomagot használom.

RNS tájképek tulajdonságai

Az alapkérdés, hogy mennyiben változtatja meg az evolúció menetét, ha nem a szekvenciát kell megőrizni, hanem a *szerkezetet*? Milyen a szerkezete annak a tájképnek, amelyben minden RNS szekvenciához (legyen ez a G genotípus-terünk) hozzárendeljük a legkisebb energiájú másodlagos szerkezetét (F fenotípus-tér). Azaz a $\Phi : s \in G \rightarrow \phi \in F$ genotípus – fenotípus megfeleltetés alapja a másodlagos szerkezet meghatározása. Az RNS-világ fejlődésének vizsgálata mellett egy ilyen tájkép vizsgálata azért is fontos, mert ez az egyetlen ismert, fizikai-kémiai alapokon nyugvó genotípus-fenotípus megfeleltetés.

A megfeleltetés alapvető tulajdonsága, hogy lényegesen kevesebb struktúra van, mint szekvencia. Tudjuk, a lehetséges N nukleotid hosszú szekvenciák száma 4^N . Az álcsomókkal rendelkező struktúrákat figyelmen kívül hagyva Schuster és munkatársai (1994) a lehetséges struktúrák számát $1,4848 \times N^{-3/2} (1,84892)^N$ -re teszik (a becslés hosszú RNS-ekre pontos). A kutatócsoport egy későbbi tanulmányában (Stadler és Haslinger) gráf-elméleti és kombinatorikai számítások alapján arra a következtetésre jutottak, hogy az álcsomók figyelembevételével a biológiailag lehetséges másodlagos struktúrák száma a szekvencia hosszával (N) körülbelül $2,35^N$ szerint növekszik. Mivel tehát lényegesen több szekvencia van, mint struktúra,

így egyértelműen következik, hogy különböző szekvenciák ugyanazt a másodlagos struktúrát veszik fel.

A másodlagos struktúra modell alapján kevés (1-3) mutáció általában nem változtatja meg a szerkezetet, de ennél több már biztosan. Az eredeti szekvencia kis mutációs lépésekkel olyan szekvenciává alakítható, amelynek bázissorrendje teljesen eltér az eredetitől, de struktúrája azonos (és minden közbülső szekvencia struktúrája is). Ezt a jelenséget hívjuk úgy, hogy egy út a tájképen perkolál. Itt a perkoláció azt jelenti, hogy van út a tájképe két pontja között, tehát az átjárás nem akadályozot. A tájképen az utak 21,7%-a teljesen perkolál és olyan szekvenciában végződik, aminek semmi köze az eredetihez. Ez rendkívül fontos az evolúció szempontjából, mert így egyrészt bizonyos mutációk semlegesek, másrészt a szekvenciater teljes egésze bejárható, ami segít az adaptációban (Huynen, *és mtsi.* 1996, Huynen 1996, van Nimwegen, *és mtsi.* 1999).

Az evolúció szempontjából egy másik fontos tulajdonsága a tájképnek, hogy minden szekvencia körül van egy nem túl nagy átmérőjű, körülbelül 20 bázisnyi hipergömb, amin belüli szekvenciák az összes gyakori struktúrát tartalmazzák. Ez az alapja a mesterséges *in vitro* kísérletek lehetőségének: a kiindulási véletlenszerű szekvenciák a szekvenciater egy nagyon kicsi töredékét képviselik, de közel hozzájuk minden gyakoribb struktúra, így a kívánt enzimaktivitáshoz szükséges struktúra is megtalálható. Feltehető, hogy a ribozimek szerkezete a *gyakori* szerkezetek között van. A szekvencia hosszának növekedésével ugyanis a szekvenciák legtöbbször a kevés számú *gyakori* struktúra egyikébe gombolyodik, míg ezek a szerkezetek a lehetséges szerkezetek elhanyagolhatóan kis részét képviselik. Ha egy enzimaktivitáshoz nem a gyakori struktúrák egyike szükséges, akkor megtalálásának valószínűsége lényegében nulla (a ritka struktúrákra az utak nem perkolálnak).

Mind a semleges utak (*neutral path*) léte, mind a gyakori struktúrák megtalálhatóságának ténye igen fontos az RNS-világ fejlődése szempontjából. Mivel a szekvenciater bejárható a rátermettség változása nélkül, s mivel ez minden gyakori struktúrára igaz, így minden bizonnyal a szekvenciater valamely részén ezek az utak közel futnak egymáshoz. Ha két ilyen út közel van egymáshoz, akkor egy működő ribozim mutánsai között lehet olyan, amely felveszi ezt az új szerkezetet, s ha ez egy az RNS populációban nem létező enzimaktivitást mutató szerkezetbe gombolyodik, akkor a populáció katalitikus repertoárja növekszik. Azaz új ribozimek kereshetők és találhatóak a meglévő mutánsai között. A semleges utak léte pedig lehetővé teszi,

hogy a szerkezet és így az enzimaktivitás megmaradjon annak ellenére, hogy a szekvenciainformáció egy része elveszik.

Ribozim rátermettségtájképek

A tudomány mára eljutott arra a szintre, hogy elég adat áll rendelkezésre, hogy egy rátermettségtájképet szerkeszthessünk valós ribozimokon végzett kísérletek eredményei alapján.

Ez először munkatársaimmal mi végeztük el (Kun, és *mtsi.* 2005). A célkitűzés megvalósításához ki kellett választanunk az irodalomból egy olyan ribozimet, amelynek másodlagos struktúrája ismert, álcsonót nem tartalmaz, nem túl hosszú és elegendő információ áll róla rendelkezésre a rátermettség-tájkép megszerkesztéséhez.

Bár egyre növekvő számú ribozimet ismerünk, ezek csak egy kis töredékének ismerjük a másodlagos szerkezetét. A közölt másodlagos szerkezetű ribozimek egy részében az elméleti módszerekkel jósolt másodlagos szerkezetet közlik, annak kísérletes igazolása nélkül (pl. Prudent, és *mtsi.* 1994, Illangasekare, és *mtsi.* 1995, Jones, és *mtsi.* 2001, Geyer és Sen 1997, Wecker, és *mtsi.* 1996). Egyedül a természetes ribozimeknek, ezek származékainak (pl. Jaeger, és *mtsi.* 1999, Piccirilli, és *mtsi.* 1992, Beaudry és Joyce 1992, Lehman és Joyce 1993, Tsang és Joyce 1994, Robertson és Ellington 2000, Tsang és Joyce 1996) és egyes ligázoknak ismerjük a kísérletileg bizonyított másodlagos szerkezetét. Ezen belül is csak a természetes ribozimeket vizsgálták kellő mértékben ahhoz, hogy a molekula más tulajdonságait, például az aktív helyet is ismerjük.

A természetes ribozimekből az I-es és II-es osztályú intronok és az RNáz P ribozim méretükből eredően nem jöhetnek számításba. A Hepatitis Delta Vírus, a kalapácsfej, a hajtükanyar és a *Neurospora* VS-ribozim, valamint a mesterséges I-es osztályú ligáz mérete megfelelő.

Enzimaktivitásából eredően a legelső jelölt a vizsgálatra az I-es osztályú ligáz volt. Ezzel a ribozimmal viszonylag sokat foglalkoztak (McGinness, és *mtsi.* 2002, Wright és Joyce 1997, Bartel és Szostak 1993, Ordoukhanian és Joyce 1999, Lehman és Riley 2003) és a „replikáz” ribozimot is ebből fejlesztették ki. Sajnos azonban egyrésről a másodlagos struktúrája álcsonót tartalmaz (1. ábra), másrészt a számos vizsgálat ellenére az enzim aktív helyét nem tudták azonosítani.

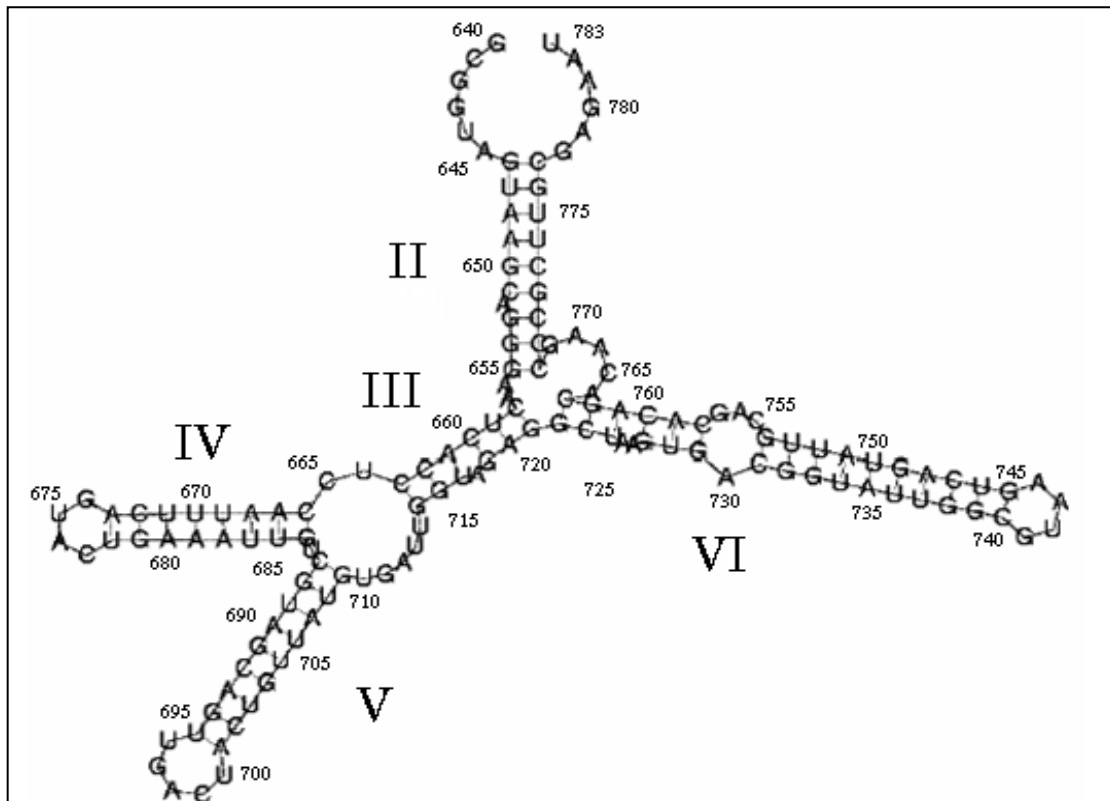
A Hepatitis Delta Vírus másodlagos struktúrája szintén álcsonót tartalmaz. A fennmaradt három jelöltből végül a *Neurospora* Varkun Szatellit ribozim és a

hajtúkanyar mellett döntöttünk. A *Neurospora* VS ribozim közepesen hosszú (164 bázis) és mutációkkal rendkívül jól karakterizált, az aktív helye és a reakcióban résztvevő egyéb pozíciói is ismertek (Lafontaine, és *mtsi.* 2002, Collins 2002, Lafontaine, és *mtsi.* 2002), így a fitnessztájkép megszerkeszthető.

***Neurospora* Varkund Szatellit ribozim**

A *Neurospora* Varkund Szatellit ribozimot (röviden VS-ribozim) először Savill és Collins (1990) izolálta *Neurospora* mitokondriumból (Collins 2002). A gomba Varkund-1c izolátumában találtak egy addig ismeretlen RNS-molekulát, amely mindig egy plazmiddal fordul elő (a *Neurospora intermedia* öt, és a *Neurospora sitophila* egy természetes izolátumában található meg ez a plazmid (Collins és Saville 1990)) (ezért hívják szatellitnek). Mint oly sokszor a tudományban az enzimaktivásra ez esetben is véletlenül derült fény. Az új ribozim a korábban felfedezett természetes ribozimek egyikére sem hasonlít, de azokhoz hasonlóan önhasítást katalizál. A ribozim a fordított reakciót, a hasított részek összekapcsolását is képes katalizálni (Saville és Collins 1991. Megjegyezendő, hogy a ligáz aktivitás tovább növelhető az 5'- és 3'-vég kiterjesztésével (Jones, Ryder és Strobel 2001).

A minimális 164 nukleotid hosszú önhasító ribozim másodlagos struktúrájának meghatározásához elméleti és kísérletes módszerek együttesét használták (Beattie, és *mtsi.* 1995). Először alternatív szerkezeteket számoltak a bevezetésben leírt algoritmussal. A jóslott legstabilabb illetve ezen struktúra energiájánál legfeljebb 10%-al kevésbé stabilabb struktúrákat vettek figyelembe. Ezt követően az eredeti szekvencia számos mutánsának mérték meg az aktivitását. Feltételezték, hogyha egy olyan nukleotidot változtatnak meg, amely egy bázispár része, akkor az aktivitás jelentősen csökken, míg ha ezt követően a feltételezett párját olyanná alakítják, hogy az újból bázispárt tudjon képezni, akkor az aktivitás visszaáll az eredeti érték közelébe. Így lehetséges az egyes feltételezett szerkezetek bázispárjainak tesztelése. Ezen eljárással végül az alternatív szerkezetek közül a 4.. ábrán látható szerkezet bizonyult a legmegfelelőbbnek. A szerkezetet összehasonlító gélelektroforézis és fluoreszcens rezonanciaenergia-transzfer (FRET) (Lafontaine, Norman és Lilley 2002, Lafontaine, és *mtsi.* 2001, Lafontaine, és *mtsi.* 2002), valamint nukleotid analógok, deléciók és kémiai módosítások (Beattie, Olive és Collins 1995, Beattie és Collins 1997, Sood, és *mtsi.* 2002, Sood és Collins 2002, Hiley, és *mtsi.* 2002) segítségével



4. ábra: A *Neurospora* Varkund Szatellit ribozim másodlagos szerkezete. A számozás megfelel az irodalomban elfogadottnak. A római számok a ribozim különböző régióit jelölik.

határozták meg. A ribozim kristályszerkezete viszont – tudtommal – röntgen-diffrakcióval a mai napig nem lett meghatározva.

A hat elkülöníthető részből álló ribozim (4. ábra) I-es részében található a hasítási hely, ez a 20 nukleotid tekinthető a szubsztrátnak. Az I-es részt lehasítva a ribozim többi részéről, a ribozim *transz*-hasításra képes (Guo és Collins 1995), így igazi enzimmént viselkedik. Más hasító enzimekkel ellentétben a VS-ribozim esetében a szubsztrátkötéshez nem szükséges számos, a ribozimmal kialakított bázispár a hasítási hely mindkét oldalán. Ennek ellenére a szubsztrát kötése igen erős (Guo és Collins 1995, Lafontaine, és *mtsi.* 2001). A szubsztrát (I-es rész) és a ribozim V-ös részének hajtókanyarjai közötti álcsumó fontos a szubsztrát orientációjában (Andersen és Collins 2001, Rastogi, és *mtsi.* 1996). A ribozim ezen részének a reakció közben felvett szerkezetét NMR-spektroszkópiával is tanulmányozták (Hoffmann, és *mtsi.* 2003).

A rátermettség-tájkép

A VS-ribozim szerkezete, a reakció mechanizmusa kellő mértékben vizsgált ahhoz, hogy a rátermettség-tájkép megszerkesztése könnyen megtehető legyen. A 144

pozícióból 87 helyen vizsgálták a mutáció hatását az enzimaktivásra. A maradék helyek jelentős része kevésbé fontos pozíciókat jelent. A mutációkkal le nem fedett részek három nagy régiót érintenek. A II-es szár elejét a 650-es pozíció előtt; a IV-es szárat, és a VI-os szár végét. A IV-es szárról és VI-os szár végéről tudjuk, hogy hosszuk jelentősen csökkenthető. A VI-os szár végén 4-6 bázispár deléciója (736-739 pozíciók) nem csökkenti az enzimaktivitást, így feltehetőleg egyedi megváltozásuk sem okoz jelentős változást. A IV-es szárban 4 bázispár kivágása (669-772) az enzimaktivitást 0,6-ra csökkenti, ami szintén nem jelentős változás. Az II-es szárról viszonylag keveset tudunk. A 652-es adenin kitüremkedés utáni részén a mutációk jelentős enzimaktivitás csökkenést okoznak. Az előtte levő helikális régiótól és a ribozim végein levő szabad végekről keveset tudunk. Feltehető azonban, hogy ez a rész a teljes – azaz az I-es szubsztrát régiót is tartalmazó – ribozimban a szubsztrát és a ribozim közötti kapcsolat miatt van és egyéb funkciója nincs. Mindemellett jó lenne tudni, hogy milyen aktivitásváltozást okoz a IV-es és a VI-os szár végén levő hajtűhurok melletti bázispár ([673:678] és [740:745]) mutációja.

Ezenkívül a rátermettség-tájkép megszerkesztéséhez fontos tudni, hogy a többszörös mutációk hogyan hatnak az enzimaktivásra. Bár végeztek olyan kísérleteket, amelyekben több egymás melletti bázispárt változtattak meg (1. táblázat), de egyrészt kevés számú ilyen van és némely esetben egy-egy mutáns itt is hiányzik (tehát lehet, hogy a kétszeres mutánst vizsgálták, de az egyszerest nem); másrészt sem a háromszoros vagy még többszörös mutációkról nincs kísérletes adatunk, sem arról, hogy a ribozim különböző régióiban bekövetkező mutációk együttes hatása milyen. Ezen adatok hiányában feltételeztem, hogy az egyes mutánsok enzimaktivitása a többszörös mutánsokban összeszorozódik. Ez volt az egyik kiinduló alapfeltételezés a rátermettség-tájkép megszerkesztésében.

A másik alapfeltételezés, amit már a VS-ribozimon kísérletezők is kimondtak (Lafontaine, Norman és Lilley 2002), hogy amíg a megfelelő struktúra ki tud alakulni, addig az egyes bázispárok konkrét identitása nem lényeges. A struktúra mellett lényegében csak a kritikus helyeket kell külön figyelembe venni.

Nem csak az eredeti ribozim szerkezetét felvevő RNS molekula rendelkezik enzimaktivitással. A deléciós és inzerciós kísérletek alternatív – bár igen közeli – szerkezeteket tártak fel, amelyeknek az enzimaktivitása még mindig jelentős. Hasonlóképpen a hibás bázispárok is lokálisan megváltoztatják a ribozim szerkezetét,

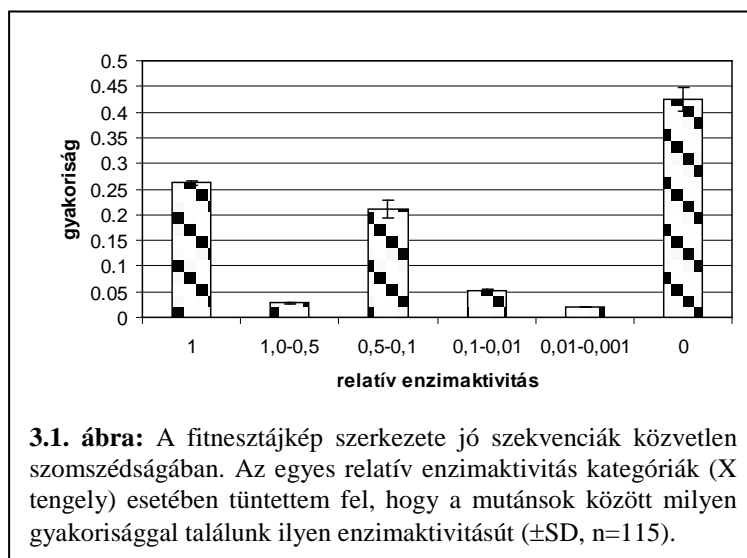
és az enzimaktivitás csökken, de az nem tűnik el. A hibás bázispárokkal így külön kell foglalkozni.

A módszerben leírt fitnessfüggvény teljesen analóg módon más ribozimokra is meghatározható. Ehhez minimálisan a ribozim másodlagos szerkezetét és a kritikus helyeket kell ismerni. Egyes alternatív szerkezetek ismerete és a hibás bázispárok járuléka nem elengedhetetlen, de megfelelően pontosítja a tájképet. Minden bizonnyal hasonlóan a hajtűhurok- és a kalapácsfej-ribozimekre is megadható a rátermettség-tájképük.

A konkrét ribozimok fitness-tájképeinek elemzése mélyebb betekintést engedhet meg az evolúciós mechanizmusokba. Érdeemes lenne mindazokat a statisztikákat és vizsgálatokat elvégezni, amelyeket a másodlagos szerkezet tájképekre elvégeztek. Szakdolgozatomban a fitness-tájkép vizsgálatának csak a kezdetét tudtam bemutatni. Az egyik legfontosabb paraméter a semleges szomszédok gyakorisága. Ez egyrészt utal a semleges utak meglétére, másrészt a tájkép csipkézettségének (*ruggedness*) mérőszáma is. Az eredmények alapján, ha egy szekvencia csak egy mutációt tartalmaz egy tökéletesen működő ribozimhoz képest, akkor 26,2% valószínűséggel a szekvencia egy tökéletes ribozimot tud kialakítani. Ez a mutációs semleges szomszédok aránya. Biztosan állítható azonban, hogy a VS-ribozimra a semleges utak nem perkolálnak abban az értelemben, hogy nem lehet az eredeti ribozim szekvenciájától teljesen eltérő, de jó ribozimot kihozni, a kritikus helyek miatt (például a 730-as helyen adeninnek kell lennie).

A leírt rátermettségfüggvény alapján megalkotható rátermettség-tájkép jellemzésére az eredeti szekvencia körüli szekvenciateret vizsgáltam. Az összes lehetséges (432 darab), az eredeti szekvenciától 1 helyen különböző mutánsból 114 (26,4%) az eredetivel azonos enzimaktivitással rendelkezik. A mutánsok 51,4%-nak van 0,1 felett az aktivitása, és 58,8%-a rendelkezik valamilyen aktivitással. A mutánsok 41,2%-a egyáltalán nem rendelkezik enzimaktivitással. Az eredeti szekvenciától két pozícióban különböző 92664 szekvencia 7,1%-a rendelkezik az eredetivel azonos enzimaktivitással. A mutánsok 31,4%-ának van enzimaktivitása, és 68,6%-nak nincs.

Az eredeti szekvencia közvetlen szomszédságában levő 114, azonos enzimaktivitással rendelkező szekvencia körül is megvizsgáltam a tájképet (3.1. ábra). A mutációk átlagosan 26,2%-a semleges (legfeljebb 27,1%, legalább 24,7% az adatok alapján).



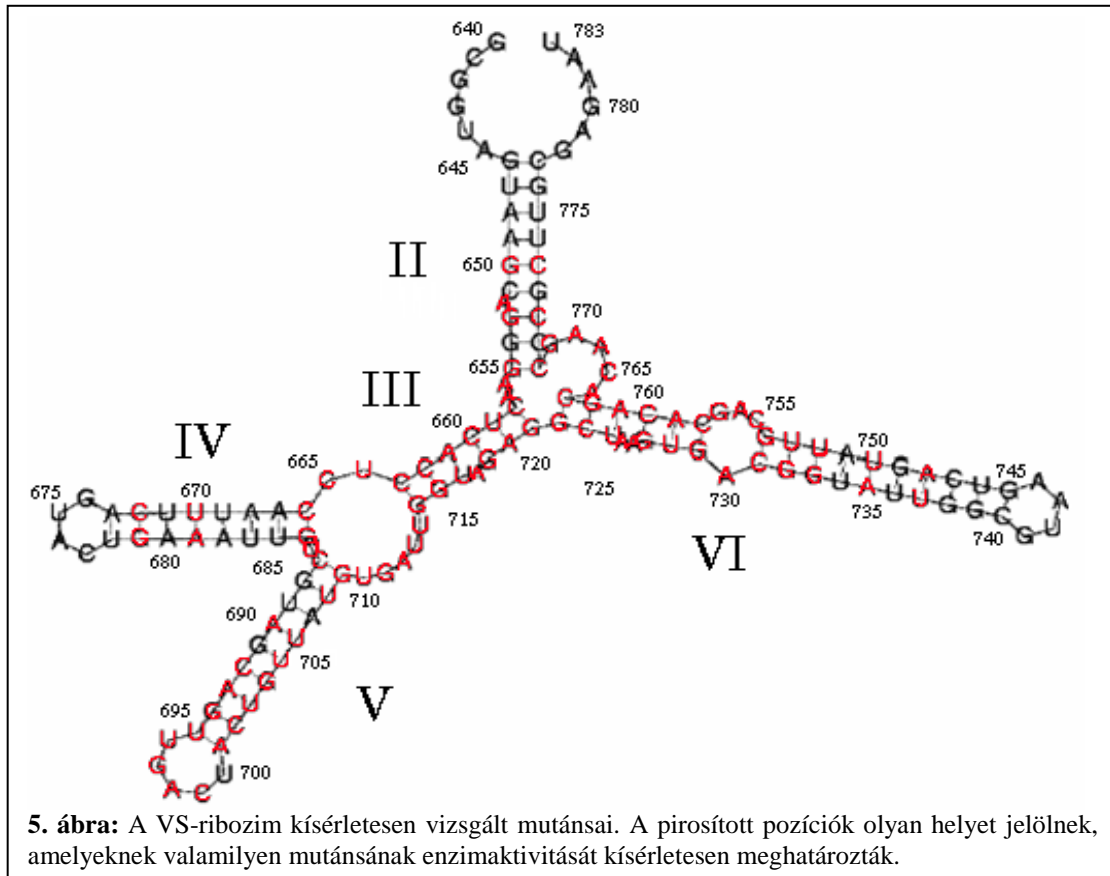
A rátermettség tájkép megalkotása

Kísérletes adatok

A rátermettség-tájkép megalkotásában azon kísérletes tanulmányok eredményei használhatóak, amelyekben különböző mutánsok (a mutáció lehet pontmutáció, inzerció vagy delécio) aktivitását határozták meg. A kísérletek célja a ribozim aktív helyének, illetve az aktivitásban résztvevő más részeinek meghatározása volt, azonban ezek az eredmények jól alkalmazhatók egy rátermettség-tájkép megalkotásához.

A VS-ribozim esetében igen szép számú mutáns szekvencia katalitikus aktivitását ismerjük. A szorosabb értelemben vett ribozim 144 nukleotidja (a II-V rész, a bevett számozást követve a 640-783 pozíciók közötti rész) közül 87 pozíción dokumentáltak valamilyen mutációt (deléciokat és inzerciókat nem számítva) (5. ábra). Az összes vizsgált mutánsok száma 183*.

* Csak a 640-783 pozíciók által definiált részt érintő mutációkat vettem figyelembe. Nem vettem figyelembe azokat a mutánsokat, amelyekben egy-egy nukleotid valamilyen nem természetes nukleotidra van cserélve.



A kísérletek alapján elmondható, hogy a ribozim másodlagos struktúrája fontos, de a bázisok legtöbbször természetük nem, mindaddig, amíg a megfelelő struktúra ki tud alakulni (Lafontaine, Norman és Lilley 2002). Egyes helyeken azonban az adott nukleotid kémiai identitása is fontos, ezeket hívjuk kritikus helyeknek (lásd Lafontaine, Norman és Lilley 2002 közleményének 6. ábráját). A másodlagos struktúra, a konkrét szekvencia és az aktivitások összefüggéséről levonható tapasztalatokat és tényeket a következőkben foglalom össze.

Bázispárok

Az egyes bázispárok legtöbbször a konkrét kémia természetük nem fontos. Többjük az aktivitás jelentős csökkenése nélkül megfordítható vagy más bázispárra cserélhető, mindaddig, amíg a bázispár ki tud alakulni az adott pozícióban.

A dokumentált 31 bázispár-cseréből az elméleti jóslás alapján 3 esetben változott meg a struktúra. Mindezek a bázispárok a VI-os száron a 2-3-6 elágazás és az aktív hely (lásd lentebb) között helyezkednek el. A további 28 mutánsból 1 aktivitása kisebb, mint 0,01; 5 aktivitása esik 0,01-0,1 közé; 7 aktivitása esik 0,1-0,75 közé; a maradék 17 esetben a relatív aktivitás 0,75 fölötti. Mivel az aktivitás meghatározásának hibája 10-25% közöttire becsülhető, ez utóbbi esetekben elmondhatjuk, hogy a ribozim aktivitása nem változik. A változtatásra jelentős

aktivitáscsökkenéssel reagáló bázispárok zöme kitüremkedés, elágazás vagy az aktív hely szomszédságában van.

Szárak hossza

A II-es szár méretét egyetlen mutációval sem változtatták, erre nézve adattal nem rendelkezünk. A III-as szár mérete nem változtatható (Lafontaine, Norman és Lilley 2002). A IV-es szár lényegében eldobható, de a 3-4-5 elágazásnak ki kell alakulnia (bár a teljes IV és V rész eltávolítása után is marad enzimaktivitása a ribozimnek (Sood és Collins 2002)). Az V-ös hélix hossza szintén kevésbé módosítható (Lafontaine, Norman és Lilley 2002). A VI-os hélix vége nagyobb aktivitásvesztés nélkül eltávolítható.

Kitüremkedő bázisok

A II, III és VI részben találunk kitüremkedő adenineket (652, 718 és 725, 726 pozíciókon). A kitüremkedések eltávolítása jelentősen csökkenti az enzimaktivitást, a másik szátra helyezésük még inkább. Más nukleotidra cserélésük azonban nem okoz jelentős változást. A III-as részben levő A718 párosítható (a 660-as pozíció után beillesztett uracillal) az aktivás szignifikáns változása nélkül (Beattie, Olive és Collins 1995, Lafontaine, Norman és Lilley 2002).

Elágazások

A ribozim két elágazása (3-4-5 és a 2-3-6) fontos szerepet játszhat a ribozim megfelelő struktúrájának kialakításában (Lafontaine, Norman és Lilley 2001, Lafontaine, Norman és Lilley 2002). Az elágazásokban egy pár helyen (kritikus helyek: 656, 657, 665, 686, 710, 712, 713, 767 és 768 pozíció) csak bizonyos nukleotid fordulhat elő, de a legtöbb pozícióban a nukleotid másra cserélhető az aktivitás jelentős csökkenése nélkül.

Álhurok helye

A szubsztrát az V-ös szár végén levő hajtűhurok 697-699 bázisaival kialakított bázispárokkal kapcsolódik a ribozimhez. Ezek a helyek így kritikus helynek tekinthetők.

Az A730-as belső hurok

A VI-os részben levő belső hurok (730, 755, 756 és 757 pozíciók) mutatta legnagyobb érzékenységet a nukleotidcserékre (Sood és Collins 2002, Lafontaine, Wilson, Norman és Lilley 2001). A szekvencia bármely megváltoztatása esetén jelentősen csökkent az enzimaktivitás. Minden bizonnyal ez az aktív helye a

ribozimnak. A hurkot alkotó négy nukleotid közül az A756 tűnik a legfontosabbnak, ezen belül is a purin 6-os pozícióján levő aminocsoport (Lafontaine, és mtsi. 2002).

Többszörösen mutált szekvenciák aktivitása

A legtöbb mutáció vagy egy bázispárt, vagy egy bázispárban részt nem vevő nukleotidot érintett. Csekély számú olyan mutáns aktivitása ismert, amely több egymástól független pozíciót is érint. Ezek a mutánsok azért fontosak, mert a fitness tájkép megszerkesztéséhez meg kell határoznunk a többszörös mutánsok aktivitását is, azaz az olyan mutánsokét, amelyek számos helyen eltérnek az eredeti szekvenciától.

A legegyszerűbb azt feltételeznünk, hogy ha egy csak m_1 helyen mutált szekvencia aktivitása a_1 , és egy tőle független m_2 helyen mutált szekvencia aktivitása a_2 , akkor az m_1 és m_2 mutációval együttesen rendelkező mutáns aktivitása $a_{mutáns} = a_1 \cdot a_2$. A fenti feltételezést támasztja alá, hogy kísérletesen ilyen összefüggést kaptak fehérjére (Tekada, és mtsi. 1989), valamint ez az egyszerű szabály jól becsli a VS-ribozim meglevő többszörös mutánsainak aktivitását (1. táblázat).

1. táblázat: Többszörös mutánsok aktivitása az egyszeres mutánsok aktivitása alapján

Mutáns 1	Aktivitás	Mutáns 2	Aktivitás	Kétszeres mutáns mért aktivitása	Kétszeres mutáns jóslott aktivitása
U752C	0,80	U753C	0,42	0,52	0,336
A661U,U717A	0,58	C662G,G716C	0,58	0,08	0,336
G722C,C763G	0,81	C723G,G762C	0,84	0,75	0,680
G727C,C760G	0,35	U728A,A759U	1,3	0,94	0,455
G716C	0,21	U717A	0,21 ¹	0,02	0,044
C662G	0,23	A661U	0,23 ¹	0,06	0,053

¹ Mért adat nem áll rendelkezésre. Feltételeztem, hogy a mutáns 2 aktivitása megegyezik a mutáns 1 aktivitásával.

A fenti táblázatból látszik, hogy a többszörös mutánsok aktivitásának számolására bevezetett egyszerű szabály alkalmazható, bár az ismert mutánsok száma miatt annak jóságát megítélni nehéz.

A rátermettség tájkép

Az előbbi általános megállapításokra és a konkrét mutációk eredményeire támaszkodva algoritmust mutatok be, amellyel bármely 144 hosszúságú szekvenciára megadható, hogy milyen enzimaktivitása lenne a VS-ribozim által katalizált reakcióban. Az algoritmussal a 4^{144} lehetséges szekvencia mindegyikéhez rendelhető relatív enzimaktivitás, ami egyben a molekula rátermettsége is.

Az algoritmus négy lépésből áll: lehetséges struktúra, hibás bázispárok, kritikus helyek és jóslott struktúra. Minden lépésben az adott lépésnek megfelelő relatív enzimaktivitást számoljuk. A molekula relatív enzimaktivitása, és így rátermettsége is ezen részaktivitások szorzata (azaz $a_{szekvencia} = a_{struktúra} \cdot a_{hibás_pár} \cdot a_{kritikus} \cdot a_{energia}$, lásd később).

Lehetséges struktúra

Az első lépésben meg kell határoznunk, hogy az adott szekvencia összeegyeztethető-e bármely szerkezettel, amely enzimaktivitással rendelkezik. Egyes delécióval vagy inzercióval előállított mutánsok aktivitása elég tekintélyes, így érdemes őket is figyelembe venni. Ennek megfelelően nem csak a 2.1. ábrán látható struktúrát, hanem pár alternatív struktúrát is figyelembe veszek. Mivel feltételeztem, hogy a ribozim minden esetben 144 nukleotid hosszú, így a deléciók és inzerciók következtében csökkent vagy megnövekedett ribozimhosszat korrigálok: a ribozim végeiről megfelelő számú nukleotidot levágok vagy a végekre illesztetek (természetesen minden lehetséges variációban). Feltételezem, hogy ezek a nukleotidok nem vesznek részt bázispár képzésében. A használt struktúrák zárójelezett reprezentációja a B függelékben található.

Az adott szekvenciáról megállapítandó, hogy mely struktúrával kompatibilis. Egy szekvenciát egy struktúrával kompatibilisnek nevezünk, amennyiben a struktúra minden (i, j) bázispárjára igaz, hogy a szekvencia i . és j . pozícióján levő nukleotidok bázispárt tudnak képezni (azaz a A:U; U:A, G:C; C:G; U:G; G:U párok egyikét).

A struktúrában engedélyezhető bizonyos számú hibás bázispár, azaz a szekvencia működő ribozim lehet, ha a hibás bázispárookra teljesülnek a következő kritériumok*:

- *Két hibás bázispár nem lehet szomszédos.* Van 6 kísérletesen előállított mutáns, amely szomszédos helyen tartalmaz hibás bázispárt. Ebből 4 esetben ([G727C, U728A]; [G722C, C723G]; [G762C, C763G]; [A759U, C760G]) az aktivitás teljesen megszűnik. A maradék két mutáns ([G716C, U717A]; [A661U, C662G]) relatív enzimaktivitása 0,02 és 0,06, ebben sorrendben.
- *A II. számban levő 653:771; 654:770 és 655:769 bázispároknak párt kell alkotniuk.* A régióban előforduló hibás bázispár mutánsok ([G653C], [C771G],

* Az egyes mutánsokban a pozíciószámok az eredeti szekvenciában levő pozíciókra utalnak. Az alternatív struktúráknál ezek a pozíciók az eltolásoknak megfelelően változnak.

[G655C]) jobbára nem rendelkeznek enzimaktivitással, kivéve a [C769G] mutáns, amelynek a relatív enzimaktivitása 0,18.

- *Az elágazások mellett levő bázispároknak párt kell alkotniuk.* Az enzimaktivitás szempontjából az elágazások létrejötte kritikus. Ha elágazások mellett bázispárok nem alkotnának párt, akkor az elágazás struktúrája jelentősen változna, így az enzimaktivitás elveszik. Ennek megfelelően a következő bázispároknak kell párt alkotniuk: 658:721; 663:715; 666:685; 687:709; 722:763.
- *Egyik szárban sem lehet több, mint két hibás pár.* Általában egy száron belül két hibás pár jelentősen csökkenti az enzimaktivitást (0,78 [G704C, U706A]; 0,54 [U670A, C672G]; 0,28 [A748U, U750A]; 0,25 [A735U, U737]; 0,07 [A690U, C692G]; 0,06 [A661U, C662G] és 0,02 [G716C, U717A]), vagy teljesen meg is szünteti azt (lásd a fent említett szomszédos hibás párokat vagy a [G679C, A681U] mutáns). Feltételezem, hogy egy száron belül kettőnél több hibás pár esetén a molekulának nem lenne enzimaktivitása.

Amennyiben a szekvencia az adott hibás párok figyelembevételével sem kompatibilis egyetlen struktúrával sem, úgy enzimaktivitása nincs és rátermettsége 0.

Ebben a lépésben az enzimaktivitás-tényező ($a_{struktúra}$) annak a struktúrának az aktivitása, amellyel a szekvencia kompatibilis. Az egyes struktúrákhoz tartozó enzimaktivitásokat a 2. táblázat tartalmazza.

2. táblázat: Struktúrák enzimaktivitása

Struktúra	$a_{struktúra}$
Eredeti struktúra	1,000
Az A652 kitüremkedés deléciója	0,013
Az A718 kitüremkedés deléciója	0,136
Az A718 kitüremkedés párosítása	0,820
8 bázispár hosszú III. szár	0,087
7 bázispár hosszú V. szár	0,045
6 bázispár hosszú IV. szár	0,470
4 bázispár hosszú IV. szár	0,590
8 bázispár hosszú VI. szár	0,970
6 bázispár hosszú VI. szár	1,000

Az esetleges hibás párokból eredő aktivitáscsökkenést a következő lépésben vesszük figyelembe.

Hibás bázispárok

Ha egy szekvencia hibás bázispár nélkül kompatibilis a fenti szerkezetek egyikével, akkor az ebben a lépésben számolandó enzimaktivitása $a_{hibás_pár} = 1$. Minden hibás bázispár, amely a fenti szabályok alapján megengedett, csökkenti az enzimaktivitást. Minden hibás bázispárhoz tartozik egy relatív enzimaktivitás ($a_{hibás_pár,i}$), és a hibás bázispárokból eredő enzimaktivitás ezen egyedi aktivitások szorzata:

$$a_{hibás_pár} = \prod a_{hibás_pár,i}.$$

Négy különböző hibás párt különböztetünk meg:

- *Aktív hellyel szomszédos bázispárok.* Az aktív hellyel szomszédos hibás bázispárok (azaz a 731:754 és/vagy 729:758 pár hibája) az aktivitást 0,025-re csökkentik ($a_{hibás_pár,i} = 0,025$). Ezeket a bázispárokat behatóan tanulmányozták és a 6 dokumentált hibás pár alapján a mutánsok átlagos relatív enzimaktivitása $0,025 \pm 0,011$.
- *Az V-ös szár végi hajtűhurok melletti hibás bázispár* (azaz a 695:701 bázispár) az aktivitást 0,05-re csökkenti ($a_{hibás_pár,i} = 0,05$). A [U695G] mutáns enzimaktivitása 0,06 és a [A701C] mutánsé 0,04.
- *Száron belüli hibás pár* az aktivitást 0,2-re csökkenti ($a_{hibás_pár,i} = 0,2$). A kísérletesen tesztelt hibás párok, amelyek nem tartoznak az előző kategóriák egyikében sem a következő relatív enzimaktivitásokkal rendelkeznek: 0,29 [C773G]; 0,23 [C662G]; 0,21 [G716C]; 0,12 [G650C]; 0,05 [A720U]; 0,00 [U659A]. Ezen mutánsok átlagos enzimaktivitása $0,15 \pm 0,095$. Meg kell azonban jegyeznünk, hogy ezek a hibás párok a ribozim fontosabb helyeit érintik. Nincs adatunk a ribozim kevésbé fontos helyeit érintő hibás párok aktivitásával kapcsolatban, ezért valamivel kisebb aktivitás csökkenést feltételeztem.
- *A IV és VI szárok végén levő hajtűhurok melletti hibás bázispár* valószínűleg nem befolyásolja jelentősen az enzimaktivitást. Feltételeztem, hogy ezen hibás párok esetében $a_{hibás_pár,i} = 0,8$.

Kritikus helyek

A lehetséges struktúra és az esetleges hibás bázispárokból eredő aktivitáscsökkenés meghatározását követően a kritikus helyeken levő nukleotidokból eredő esetleges enzimaktivitás-csökkenést kell figyelembe vennünk.

A kritikus helyeket megfelelően vizsgálták ahhoz, hogy – bizonyos feltevésekkel – minden nukleotid jelenléte esetén meg lehessen adni az enzimaktivitást (6. táblázat). Az egyes helyek enzimaktivitásainak ($a_{kritikus,i}$) szorzata adja a kritikus helyekből eredő enzimaktivitás tagot ($a_{kritikus}$).

3. táblázat: Kritikus helyek relatív enzimaktivitása a nukleotidok függvényében

Kritikus hely	$a_{kritikus,i}$			
	A	U	C	G
656	1,000	0,002	0,003	0,061
657	1,000	0,063	0,063 ¹	0,063 ¹
665	0,014	0,01 ²	1,000	0,01
686	0,006	1,000	0,006 ¹	0,006 ¹
697	0,000	0,001	0,000	1,000
698	1,000	0,000	0,041	0,000
699	0,006	0,018	1,000	0,000
710	0,002	1,000	0,029	0,002
712	1,000	0,005	0,005 ²	0,006
713	0,019	1,000	0,136	0,019
730	1,000	0,036	0,055	0,011
755	0,840	0,170	1,000	0,019
756	1,000	0,001	0,002	0,002
757	0,044	0,018	0,014	1,000
767	1,000	0,047	0,047 ¹	0,047 ¹
768	0,573	0,015	0,573 ³	1,000

¹ Adat nem áll rendelkezésre. Feltételezem, hogy az aktivitás az ismert mutáns enzimaktivitásával egyenlő.

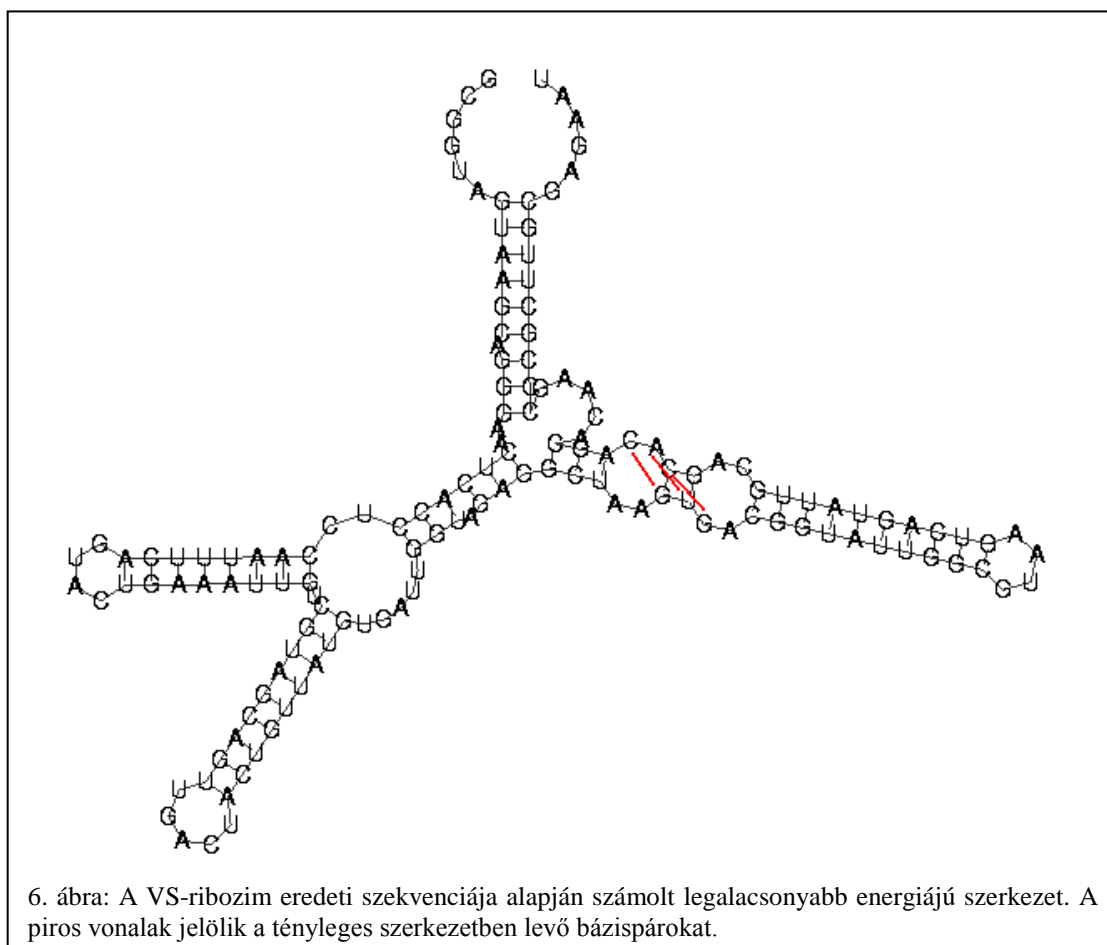
² Adat nem áll rendelkezésre. Feltételezem, hogy az aktivitás az ismert mutánsok közül a kisebb enzimaktivitású enzimaktivitásával egyenlő.

³ Adat nem áll rendelkezésre. Feltételezem, hogy az aktivitás az ismert mutánsok közül a nagyobb enzimaktivitású enzimaktivitásával egyenlő.

Jóslott struktúra

A szekvencia rátermettségének meghatározásának technikai okok miatt az utolsó lépés az, hogy elméleti számolással meghatározzam a szekvencia struktúráját és ezt összevessem a kívánt struktúrával. Ez a lépés a legutolsó, mivel ez a legidőigényesebb, így csak olyan szekvenciákra végzem el, amelyek rátermettsége az eddigiek alapján $a_{struktúra} \cdot a_{hibás_pár} \cdot a_{kritikus} \geq 0,001$. A 0,001 általában az enzimaktivitás kimutatásának alsó határa, az ennél kisebb enzimaktivitású ribozimeket nem tudják elkülöníteni az aktivitással egyáltalán nem rendelkezőektől. Ezért az ilyen alacsony enzimaktivitású ribozimek rátermettségét 0-nak vettem.

A VS-ribozim eredeti szekvenciájának a *Vienna Package* segítségével megállapított legalacsonyabb energiájú másodlagos struktúrája (6. ábra) nem azonos a kísérletesen



megállapított másodlagos struktúrával. Az eltérés oka több minden lehet, legvalószínűbb, hogy az energiaparaméterek nem tökéletesek a ribozimhez. Mivel azonban a két struktúra között az energiakülönbség összesen 0,3 kcal, és 3 bázispárban térnek el egymástól, így az eljárást nyugodtan használhatjuk a rátermettség megállapításához.

A másodlagos struktúra számolása előtt a várt másodlagos struktúrát meg kell szerkeszteni. A várt másodlagos struktúra egyrészt az első lépésben meghatározott szerkezettől, másrészt a jelenlevő hibás pároktól függ. A hibás párokból eredő rátermettség-csökkenést már figyelembe vettük, így a másodlagos szerkezetben jelenlétüket megengedjük.

A jóslt és a tényleges struktúra különbözősége következtében az adott szekvenciához két várt struktúra is tartozik, az egyik a tényleges másodlagos struktúra, a másik a számoltnak megfelelő (tehát 728-729 régióban eltérő) másodlagos

struktúra. A szekvenciára számolt legkisebb energiájú struktúrának valamelyik várt struktúrával meg kell egyeznie. Egyezés esetén az $a_{energia} = 1$, különben $a_{energia} = 0$.

Mindkét struktúra figyelembevételére azért van szükség, mert bár egyes mutánsoknál az algoritmus jól jósolja a másodlagos struktúrát, a zöménél nem.

A fenti négy lépésben számolt enzimaktivitás tagok szorzata adja meg az adott molekula becsült relatív enzimaktivitását.

$$a_{szekvenca} = a_{struktura} \cdot a_{hibas_par} \cdot a_{kritikus} \cdot a_{energia} \quad (3)$$

Felhasznált irodalom

- P. G. Higgs. 2000. RNA secondary structure: physical and computational aspects. - *Quarterly Reviews of Biophysics* **33**: 199-253.
- R. B. Lyngsø és C. N. S. Pedersen. 2000. RNA pseudoknot prediction in energy-based models. - *Journal of Computational Biology* **7**: 409-427.
- H. Isambert és E. D. Siggia. 2000. Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. - *Proceedings of the National Academy of Sciences of the USA* **97**: 6515-6520.
- D. H. Mathews, J. Sabina, M. Zucker és H. Turner. 1999. Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. - *Journal of Molecular Biology* **288**: 911-940.
- M. S. Waterman. 1978. Secondary structure of single – stranded nucleic acids. - *Studies on foundations and combinatorics, Advances in mathematics supplementary studies, Academic Press N.Y.* **1**: 167-212.
- M. S. Waterman és T. F. Smith. 1978. RNA secondary structure: a complete mathematical analysis. - *Mathematical biosciences* **42**: 257-266.
- R. Nussinov és A. B. Jacobson. 1980. Fast algorithm for predicting the secondary structure of single-stranded RNA. - *Proceedings of the National Academy of Sciences of the USA* **77**: 6309-6313.
- R. Nussinov, G. Piecznik, J. R. Griggs és D. J. Kleitman. 1978. Algorithms for loop matching. - *SIAM Journal of Applied Mathematics* **35**: 68-82.
- M. Zuker és D. Sankoff. 1984. RNA secondary structures and their prediction. - *Bulletin of Mathematical Biology* **46**: 591-621.
- M. Zuker és P. Stiegler. 1981. Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. - *Nucleic Acid Research* **9**: 133-148.
- I. L. Hofacker, W. Fontana, P. F. Stadler, S. Bonhoeffer, M. Tacker és P. Schuster. 1994. Fast folding and comparison of RNA secondary structures. - *Monatshefte für Chemie* **125**: 167-188.
- P. Schuster, W. Fontana, P. F. Stadler és I. L. Hofacker. 1994. From sequences to shapes and back: a case study in RNA secondary structures. - *Proceedings of the Royal Society of London B* **255**: 279-284.
- P. F. Stadler és C. Haslinger. RNA structure with pseudo-knots. -
- M. A. Huynen, P. F. Stadler és W. Fontana. 1996. Smoothness within ruggedness: the role of neutrality in adaptation. - *Proceedings of the National Academy of Sciences of the USA* **93**: 397-401.
- M. A. Huynen. 1996. Exploring phenotype space through neutral evolution. - *Journal of Molecular Evolution* **43**: 165-169.
- E. van Nimwegen, J. P. Crutchfield és M. A. Huynen. 1999. Neutral evolution of mutational robustness. - *Proceedings of the National Academy of Sciences of the USA* **96**: 9716-9720.
- Á. Kun, M.-C. Maurel, M. Santos és E. Szathmáry. 2005. Fitness landscapes, error thresholds, and cofactors in aptamer evolution. - In: S. Klussmann (ed.) *The aptamer handbook*. WILEY-VCH Verlag GmbH & Co. KGaA, old. 54-92.
- J. R. Prudent, T. Uno és P. G. Schultz. 1994. Expanding the scope of RNA catalysis. - *Science* **264**: 1924-1927.

- M. Illangasekare, G. Sanchez, T. Nickles és M. Yarus. 1995. Aminoacyl-RNA synthesis catalyzed by an RNA. - *Science* **267**: 643-647.
- F. D. Jones, S. P. Ryder és S. A. Strobel. 2001. An efficient ligation reaction promoted by a Varkund satellite ribozyme with extended 5'- and 3'-termini. - *Nucleic Acid Research* **29**: 5115-5120.
- C. R. Geyer és D. Sen. 1997. Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme. - *Chemistry & Biology* **4**: 579-593.
- M. Wecker, D. Smith és L. Gold. 1996. In vitro selection of a novel catalytic RNA: characterization of a sulfur alkylation reaction and interaction with small peptide. - *RNA* **2**: 982-994.
- L. Jaeger, M. C. Wright és G. F. Joyce. 1999. A complex ligase ribozyme evolved *in vitro* from a group I ribozyme domain. - *Proceedings of the National Academy of Sciences of the USA* **96**: 14712-14717.
- J. A. Piccirilli, T. S. McConnell, A. J. Zaug, H. F. Noller és T. R. Cech. 1992. Aminoacyl esterase activity of the *Tetrahymena* ribozyme. - *Science* **256**: 1420-1424.
- A. A. Beaudry és G. F. Joyce. 1992. Directed evolution of an RNA enzyme. - *Science* **257**: 635-641.
- N. Lehman és G. F. Joyce. 1993. Evolution *in vitro* of an RNA enzyme with altered metal dependence. - *Nature* **361**: 182-185.
- J. Tsang és G. F. Joyce. 1994. Evolutionary optimization of the catalytic properties of a DNA-cleaving ribozyme. - *Biochemistry* **33**: 5966-5973.
- M. P. Robertson és A. D. Ellington. 2000. Design and optimization of effector-activated ribozyme ligases. - *Nucleic Acid Research* **28**: 1751-1759.
- J. Tsang és G. F. Joyce. 1996. Specialization of the DNA-cleaving activity of a Group I ribozyme through *in vitro* evolution. - *Journal of Molecular Biology* **262**: 31-42.
- K. E. McGinness, M. C. Wright és G. F. Joyce. 2002. Continuous *in vitro* evolution of a ribozyme that catalyzes three successive nucleotidil addition reaction. - *Chemistry & Biology* **9**: 585-596.
- M. C. Wright és G. F. Joyce. 1997. Continuous *in vitro* evolution of catalytic function. - *Science* **276**: 614-617.
- D. P. Bartel és J. W. Szostak. 1993. Isolation of a new ribozyme from a large pool of random sequences. - *Science* **261**: 1411-1418.
- P. Ordoukhanian és G. F. Joyce. 1999. A molecular description of the evolution of resistance. - *Chemistry & Biology* **6**: 881-889.
- N. Lehman és C. A. Riley. 2003. Expanded metal-ion tolerance of evolved ligase ribozymes. - *Biochimie* **85**: 683-689.
- D. A. Lafontaine, D. G. Norman és D. M. J. Lilley. 2002. Folding and catalysis by the VS ribozyme. - *Biochimie* **84**: 889-896.
- R. A. Collins. 2002. The *Neurospora* Varkund satellite ribozyme. - *Biochemical Society Transactions* **30**: 1122-1126.
- D. A. Lafontaine, D. G. Norman és D. M. J. Lilley. 2002. The structure and active site of the Varkund satellite ribozyme. - *Biochemical Society Transactions* **30**: 1170-1175.
- B. J. Saville és R. A. Collins. 1990. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. - *Cell* **61**: 685-696.
- R. A. Collins és B. J. Saville. 1990. Independent transfer of mitochondrial chromosomes and plasmids during unstable vegetative fusion in *Neurospora*. - *Nature* **345**: 719-724.
- B. J. Saville és R. A. Collins. 1991. RNA-mediated ligation of self-cleavage products of a *Neurospora* mitochondrial plasmid transcript. - *Proceedings of the National Academy of Sciences of the USA* **88**: 8826-8830.
- T. L. Beattie, J. E. Olive és R. A. Collins. 1995. A secondary-structure model for the self-cleaving region of the *Neurospora* VS RNA. - *Proceedings of the National Academy of Sciences of the USA* **92**: 4686-4690.
- D. A. Lafontaine, D. G. Norman és D. M. J. Lilley. 2001. Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction. - *The EMBO Journal* **20**: 1415-1424.
- D. A. Lafontaine, D. G. Norman és D. M. J. Lilley. 2002. The global structure of the VS ribozyme. - *The EMBO Journal* **21**: 2461-2471.
- T. L. Beattie és R. A. Collins. 1997. Identification of functional domains in the self-cleaving *Neurospora* VS ribozyme using damage selection. - *Journal of Molecular Biology* **267**: 830-840.
- V. D. Sood, S. Yekta és R. A. Collins. 2002. The contribution of 2'-hydroxyls to the cleavage activity of the *Neurospora* VS ribozyme. - *Nucleic Acid Research* **30**: 1132-1138.
- V. D. Sood és R. A. Collins. 2002. Identification of the catalytic subdomain of the VS ribozyme and evidence for remarkable sequence tolerance in the active site loop. - *Journal of Molecular Biology* **320**: 443-454.

- S. L. Hiley, V. D. Sood, J. Fan és R. A. Collins. 2002. 4-thio-U cross-linking identifies the active site of the VS ribozyme. - *The EMBO Journal* **21**: 4691-4698.
- H. C. T. Guo és R. A. Collins. 1995. Efficient *trans*-cleavage of a stem-loop RNA substrate by a ribozyme derived from *Neurospora* VS RNA. - *The EMBO Journal* **14**: 368-376.
- D. A. Lafontaine, T. J. Wilson, D. G. Norman és D. M. J. Lilley. 2001. The A730 loop is an important component of the active site of the VS ribozyme. - *Journal of Molecular Biology* **312**: 663-674.
- A. A. Andersen és R. A. Collins. 2001. Intramolecular secondary structure rearrangement by the kissing interaction of the *Neurospora* VS ribozyme. - *Proceedings of the National Academy of Sciences of the USA* **98**: 7730-7735.
- T. Rastogi, T. L. Beattie, J. E. Olive és R. A. Collins. 1996. A long-range pseudoknot is required for activity of the *Neurospora* VS ribozyme. - *The EMBO Journal* **15**: 2820-2825.
- B. Hoffmann, G. T. Mitchell, P. Gendron, F. Major, A. A. Andersen, R. A. Collins és P. Legault. 2003. NMR structure of the active conformation of the Varkund satellite ribozyme cleavage site. - *Proceedings of the National Academy of Sciences of the USA* **100**: 7003-7008.
- D. A. Lafontaine, T. J. Wilson, Z.-Y. Zhao és D. M. J. Lilley. 2002. Functional group requirements in the probable active site of the VS ribozyme. - *Journal of Molecular Biology* **323**: 23-34.
- Y. Tekada, A. Sarai és V. M. Rivera. 1989. Analysis of the sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments. - *Proceedings of the National Academy of Sciences of the USA* **86**: 439-443.