Az élet keletkezése – A genetikai kód eredete

Szakdolgozat
biológia alapszak, biológus szakirány

készítette:
MÁRKUS BENJAMIN

témavezető:
Dr. KUN ÁDÁM, Tudományos főmunkatárs
Növényrendszertani, Ökológiai és Elméleti Biológiai Tanszék

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
BIOLÓGIAI INTÉZET

Budapest, 2013

Tartalomjegyzék

I. Bevezetés ... 4
I.1. A genetikai kód ... 4
I.1.1. A genetikai kód redundáns ... 4
I.1.2. A genetikai kód vesszőmentes és egy leolvasási keretben értelmezendő ... 5
I.1.3. A genetikai kód „lötyög” ... 6
I.1.4. A genetikai kód többé-kevésbé univerzális ... 6
I.1.5. A genetikai kódban dedikált start- és stop-kodonok találhatóak ... 7
I.2. A genetikai kód megfejtése – tudománytörténeti áttekintés ... 8
I.2.1. Előzmények: a szekvenciális hipotézis, a centrális dogma és az adapter hipotézis ... 8
I.2.2. A kód tripletes természetének felfedezése ... 9
I.2.3. A kódtábla feltárása ... 9
II. A genetikai kód eredetének problematikája ... 11
II.1. A DNS-fehérje világ eredetének kérdése ... 11
II.2. Az RNS-világ ... 13
II.2.1. Az RNS-világ előtt és után ... 13
II.2.2. Az RNS-világ szerveződése ... 14
III. Az ősi genetikai kód kialakulása ... 17
III.1. A kódoló koenzim fogantyú hipotézis ... 17
III.1.2. A nukleinsav-aminosav hozzárendelés megjelenése ... 17
III.2. A sztereokémiai hipotézis ... 21
III.3. A genetikai kód kiterjeszkedése ... 22
III.3.1. A kód kiterjeszkedésének adaptív háttere ... 23
III.3.2. A korai aminosavak halmazának azonosítása ... 24
III.3.3. A koevolúciós hipotézis ... 26
IV. A sztenderd genetikai kód mintázatai és a genetikai kód evolúciója ... 31
IV.1. A sztenderd genetikai kód mintázatai ... 31
IV.2. A sztenderd genetikai kód hibaminimalizáló jellege ... 32
IV.2.1. A hibaminimalizáló jelleg foka a sztenderd genetikai kódban ... 33
IV.2.2. A kód evolúciója és a hibaminimalizáló jellegének eredetének kérdése 35
V. Összefoglalás ... 39

VI. Summary 40
VII. Hivatkozások 42
VIII. Köszönetnyilvánítás 44

I. Bevezetés

I.1. A genetikai kód

A genetikai kód egy szabályrendszer, amely az élő szervezetekben előforduló nukleotidokból képzett három bázis hosszúságú kodonokhoz (és párjukhoz, az antikodonokhoz) egyértelműen fehérjék felépítéséhez szükséges α-aminosavakat, start-jelet vagy stop-jeleket rendel. Ez az egyértelmű hozzárendelés lehetővé teszi, hogy az élő szervezetek által hasznosított polinukleotidok, nevezetesen a dezoxiribonukleinsav (DNS) és a ribonukleinsav (RNS) adott szakaszai változatos és egyedi bázissorrendjük, vagy másképp szekvenciájuk révén hasonlóképpen változatos és egyedi polipeptideket határozzanak meg. A genetikai kód tehát egy hasonlattal élve tulajdonképpen egy szótár szerepét tölti be, amely a nukleinsavak nyelvéről az aminosavak, illetve polipeptidek nyelvére képes fordítani. A genetikai kód számos, jól definiálható jellegzetességgel és tulajdonsággal rendelkezik, ezek bemutatására térek ki a következőkben.

I.1.1. A genetikai kód redundáns

Kombinatorikai szempontból a genetikai kód által leképzett szabályrendszer egy ismétléses variációnak felel meg, így látható, hogy a négy lehetséges nukleotid felhasználásával összesen 4^3 = 64 különböző, három nukleotid hosszúságú kodon alkotható. Ha figyelembe vesszük, hogy – néhány taxon kivételével – az élő szervezetek kodontáblája általában 20 különböző aminosavat kódol, akkor beláthatjuk, hogy egy aminosavat átlagosan 64/20 = 3,2 db, vagy a három stop-kodont figyelembe véve (64 – 3)/20 = 3,05 db kodon kódol. Egy további, és külön kiemelendő összetevője a genetikai kód e tulajdonságának, hogy az azonos aminosavat kódoló kodonok a kódtáblában nem szétszórva, hanem egymáshoz a lehetséges mutációk által kifeszített térben közel helyezkednek el, kettő, négy vagy hat kodonból álló blokkokat képezve (1. ábra). A kettes és a négyes blokkokban a kodonok csak az utolsó (harmadik), a hatos leucin- és arginin-blokkokban pedig az első és az utolsó pozícióban különböznek; a kodonok második nukleotidja a blokkokon belül minden esetben azonos. A genetikai kód e tulajdonságát a genetikai

kód degeneráltságának vagy redundanciájának nevezzük. Ennek a jellegzetességnek a mélyebb implikációira és az evolúciós hátterét tárgyaló elméletekre külön is kitérek a szakdolgozatnak a sztenderd genetikai kód evolúcióját tárgyaló fejezetében, itt csak a jelenség rövid bemutatására szorítkozom.

UUU [F] Phe	UCU [S] Ser	UAU [Y] Tyr	UGU [C] Cys
UUC [F] Phe	UCC [S] Ser	UAC [Y] Tyr	UGC [C] Cys
UUA [L] Leu	UCA [S] Ser	UAA [ ] Ter	UGA [ ] Ter
UUG [L] Leu	UCG [S] Ser	UAG [ ] Ter	UGG [W] Trp
CUU [L] Leu	CCU [P] Pro	CAU [H] His	CGU [R] Arg
CUC [L] Leu	CCC [P] Pro	CAC [H] His	CGC [R] Arg
CUA [L] Leu	CCA [P] Pro	CAA [Q] Gln	CGA [R] Arg
CUG [L] Leu	CCG [P] Pro	CAG [Q] Gln	CGG [R] Arg
AUU [I] Ile	ACU [T] Thr	AAU [N] Asn	AGU [S] Ser
AUC [I] Ile	ACC [T] Thr	AAC [N] Asn	AGC [S] Ser
AUA [I] Ile	ACA [T] Thr	AAA [K] Lys	AGA [R] Arg
AUG [M] Met	ACG [T] Thr	AAG [K] Lys	AGG [R] Arg
GUU [V] Val	GCU [A] Ala	GAU [D] Asp	GGU [G] Gly
GUC [V] Val	GCC [A] Ala	GAC [D] Asp	GGC [G] Gly
GUA [V] Val	GCA [A] Ala	GAA [E] Glu	GGA [G] Gly
GUG [V] Val	GCG [A] Ala	GAG [E] Glu	GGG [G] Gly

1. ábra: genetikai kód az azonos aminosavat kódoló blokkok kiemelésével. Az árnyékolás mértéke az egyes aminosavak polaritás szempontjából vett hasonlóságát tükrözi (ennek részletes tárgyalását lásd a kód optimalizáltságáról szóló fejezetben). A kép forrása: Koonin és mtsai, 2007.

I.1.2. A genetikai kód vesszőmentes és egy leolvasási keretben értelmezendő

A transzlációs apparátus egy mRNS-molekula bázissorrendjének olvasása során az egyes kodonokat kizárólag az alapján képes elkülöníteni és azonosítani, hogy három nukleotid alkot egy kodont, és így minden nukleotidhármas után új kodon kezdődik, külön jel a kodonok elválasztására nincsen. Másképp megfogalmazva azt mondhatjuk tehát, hogy a genetikai kód vesszőmentes. Ennek egy további

következménye, hogy a genetikai kód egy leolvasási keretben értelmezendő: a kezdő kodon első nukleotidja az összes utána következő kodont kijelöli egy adott leolvasási esemény során. A leolvasási keret esetleges elcsúszása, például egy inszerció vagy egy deléció következtében, jelentős következményekkel jár, ugyanis ebben az esetben a transzláció eredménye egy működésképtelen fehérjemolekula.

I.1.3. A genetikai kód „lötyög”

Noha ez a tulajdonság nem a szigorú értelemben vett kód, hanem inkább az azt implementáló transzlációs apparátus sajátja, mégis érdemes megemlíteni, ugyanis lényegileg érinti a kódolás mechanizmusát. A kód „lötyögése”, melynek felfedezése Francis Crick nevéhez fűződik, azt a jelenséget takarja, hogy a transzláció során bizonyos esetekben egy adott mRNS kodonhoz több, különböző antikodont hordozó tRNS is képes kapcsolódni, melyek antikodonjai az első pozícióban különböznek (sok esetben itt inozitolt találunk, ez ugyanis nem kanonikus bázispárosodást képes kialakítani adenozinnal, citozinnal és uracillal is). Az, hogy ez nem eredményez hibás fehérjéket, éppen a fentebb bemutatott redundáns jellegnek köszönhető, az antikodon első pozíciója ugyanis a kodon harmadik pozíciójának felel meg, amely, mint láttuk, a blokkokon belül nem okoz aminosav-változást.

I.1.4. A genetikai kód többé-kevésbé univerzális

A genetikai kód univerzális jellege azt jelenti, hogy nagyon kevés kivételtől eltekintve a kódtábla az élővilág egészében teljesen azonos. Az általános szabály alóli kivételek, melyek főként mitokondriumoknál találhatóak, nem számosak: az amerikai Nemzeti Egészségügyi Intézet vonatkozó, 2013-ben frissített adatbázisa 19 különböző kódtáblát sorol fel (NCBI, 2013). Fontos azonban megjegyezni, hogy ezek kivétel nélkül a sztenderd kód leszármazottainak tekinthetőek, így a genetikai kód univerzális jellege egy erőteljes bizonyíték amellett, hogy a jelenlegi ismert élővilág monofiletikus, vagyis egyetlen közös őstől származik, és ez a közös ős már rendelkezett a sztenderd genetikai kóddal. A genetikai kód univerzalitása történeti nézőpontba helyezve továbbá arra utal, hogy a kód erősen konzervált, hiszen a legutolsó közös ős óta, vagyis nagyjából az elmúlt 3,5 milliárd év során kevéssé és ritkán változott. Ennek egyik első (ám közel sem egyetlen) lehetséges magyarázatát

Crick hozta a „befagyott véletlen” elmélettel (*frozen accident theory*): eszerint azzal párhuzamosan, hogy az élő szervezetekben a transzlációs apparátus alkalmazása egyre széleskörűbbé és beágyazottabbá vált, drasztikusan megnőtt a kód megváltozásának költsége, ugyanis akár csak egy darab kodonkiosztás megváltozása is számos fehérje esetén jelentene funkcióvesztést, ez pedig jó eséllyel letális. Ennek következménye, hogy a kód egy bizonyos „véletlen” állapotban „befagyott”, vagyis evolúciója lényegében megállt. A befagyott véletlen elmélettel összecsengő, érdekesnek tűnő megfigyelésem, hogy a sztenderd genetikai kódtól való eltérések döntő többségükben (a fentebb hivatkozott adatbázisban a 18 kódból 11 esetben) mitokondriumokban alakultak ki, melyek egyrészt másodlagosan lényeges mennyiségi redukáltságot mutató saját fehérjeállománnyal rendelkeznek, másrészt endoszimbionta voltukból fakadóan környezetük, a gazdasejt citoplazmája, meglehetősen konstans és a kölcsönös függésből adódóan támogató jellegű. Mindezen körülmények véleményem szerint az esetlegesen létrejövő hibás fehérjék hatásának súlyosságát tompíthatják, és ez lehetőséget nyújthat a befagyott kód részleges „felolvadására”, vagyis arra, hogy egyes kodonkiosztások megváltozzanak anélkül, hogy ez a mitokondrium funkcionális működését ellehetetlenítené.

I.1.5. A genetikai kódban dedikált start- és stop-kodonok találhatóak

A fehérjék szintézise során a sejt transzlációs apparátusa azokat a genetikai kód implementálásával, egy adott mRNS molekula szekvenciális leolvasása révén állítja elő. Mivel azonban egy mRNS molekulának mind az 5’, mind a 3’ vége tartalmaz fehérjévé át nem fordítandó szakaszokat, ezért szükséges a transzlációs apparátus számára valamiképpen kijelölni a fehérjeszintézis során ténylegesen felhasználandó mRNS-részletet. A kezdőpontot egy start-kodon jelzi: ez prokariótákban egy purin-gazdag részletben található, míg eukariótákban egyszerűen az mRNS 5’-végéhez legközelebb eső AUG kodon. A szintézis végét és a fehérje elkészültét ugyancsak speciális kodonok jelzik, ezek a stop-kodonok: az UAG, az UAA és az UGA. Figyelemre méltó, hogy míg az AUG csak speciális kontextusban funkcionál start-kodonként, egyébként pedig metionint kódol, a stop-kodonok minden esetben a transzláció terminációját eredményezik, aminosav-kódoló szerepük nincsen.

I.2. A genetikai kód megfejtése – tudománytörténeti áttekintés

I.2.1. Előzmények: a szekvenciális hipotézis, a centrális dogma és az adapter hipotézis

A genetikai kód felismerését, mint egyértelmű hozzárendelést nukleinsavak és aminosavak között, nagymértékben előkészítették Francis Crick 1958-ban közzétett, meglehetősen spekulatív fejtegetései (Crick, 1958). Fontos érdeme Cricknek, hogy ebben a publikációban az akkoriban széles körben elérhető, ám sok tekintetben addig konfúznak tetsző kísérletes eredményeket sikerrel ötvözte és vázolt fel egy koherens, a későbbi kísérletes munkák során döntően helyesnek bizonyuló modellt. A Crick által felhasznált eredmények közül a legfontosabbak a következőek voltak: 1. a „genetikai anyag" legtartósabb része DNS-ből áll, és a fehérjék valamiképpen a „genetikai anyag" szabályozása alatt állnak; 2. a „genetikai anyag" szekvenciális természetű, csakúgy, mint a fehérjék; 3. a fehérjeszintézis riboszómákon (a kor nómenklatúrája szerint: mikroszómális részecskéken) történik, melyek ribonukleoproteinből állnak; 4. a fehérjeszintézis egy adott szakaszában az aminosavakat valamilyen RNS (akkoriban: soluble, vagyis „oldható" RNS) szállítja a riboszómához. Ezen információk alapján aztán három fő megállapítást tett, melyek alapvetőnek bizonyultak a genetikai kód ötletének kidolgozásában. A szekvenciális hipotézisben Crick azt mondta ki, hogy a nukleinsavak specificitása kizárólag bázissorrendjükben rejlik, és hogy ez a specifikus bázissorrend valamiképpen a fehérjék aminosavsorrendjét is kódolja. A centrális dogma megállapításai a nukleinsavak és a fehérjék közti információátvitel irányára és lehetséges módjaira vonatkoznak: eszerint nukleinsavról nukleinsavra és fehérjére lehetséges az információ áramlása, fehérjéről nukleinsavra azonban nem. A harmadik elmélet, az adapter hipotézis pedig a közvetlen nukleinsav templát-aminosav kapcsolat fiziko-kémiai implauzibilitása alapján állapítja meg, hogy léteznie kell egy adapter molekulának, amely fehérjeszintéziskor a templát nukleinsav információtartalma alapján a megfelelő aminosavat adja a rendszerhez. Fontos megemlíteni, hogy ekkor még nem vált el egymástól az mRNS és az rRNS ideája, Crick úgy vélte, hogy a „mikroszómális részecskék" ribonukleoproteinjának ribonukleinsav részei egy az egyben a fehérjék kódolásáért felelős nukleinsav-templátot adják, azt pedig, hogy a

megjósolt adapter a nem sokkal korábban felfedezett, aminosav-szállító szerepű RNS-sel azonos lenne, csak, mint egy lehetséges alternatívát jelölte meg.

I.2.2. A kód tripletes természetének felfedezése

A genetikai kód megfejtéséhez vezető út elengedhetetlen mozzanatát jelentette a kód tripletes tagolódásának feltárása. Az ezt tisztázó 1961-es kísérlet során, mely Francis Crick, Leslie Barnett, Sydney Brenner és R. J. Watts-Tobin nevéhez fűződik, T4 bakteriofágok fertőzőképességének változását vizsgálták olyan ágensek alkalmazása mellett, amelyek inszerciókat vagy deléciókat okoznak. Az eredmények azt mutatták, hogy az egy vagy két bázis nagyságú inszerciók és deléciók a T4 fág fertőzőképességéért felelős valamelyik gén teljes funkcióvesztését okozták, három bázis beépülése vagy kiesése azonban a vad fenotípushoz nagyon hasonló, de azzal nem teljesen megegyező fertőzőképességű mutánsokat eredményez. Ezek alapján Crick és munkatársai felismerték, hogy a kód egy leolvasási keretben értelmeződik, és az ennek eltolódását eredményező mutációk állnak a fertőzőképesség elvesztése mögött. Továbbá, mivel látható volt, hogy éppen hárombázisnyi eltérések esetén nem érvényesül a leolvasási keret eltolódásának hatása, logikus volt a következtetés, hogy a kód tripletes, hiszen ekkor a leolvasási keret helyreállt. A vad fenotípustól való kis eltérést ebben az esetben a beépülő vagy kieső aminosavval magyarázták.

I.2.3. A kódtábla feltárása

Az áttörést az első kodonkiosztás megfejtésével Marshall Nirenberg és Heinrich Matthaei 1961-es kísérlete hozta. Ennek során azt találták, hogy amennyiben sejtkivonathoz csak uracilból álló homopoliribonukleotidokat (poliU-t) adnak, akkor kizárólag fenilalanin egységekből álló homopolipeptid szintetizálódik a rendszerben. A kód párhuzamosan felfedezett tripletes voltát figyelembe véve adódott a következtetés: az UUU kodon fenilalanin beépítésére utasítja a fehérjeszintézis apparátusát. Ezen a vonalon továbbmenve aztán hamarosan megfejtésre került az AAA és a CCC kodonok jelentése is (a GGG kodon megfejtése technikai problémák miatt nem volt lehetséges ezzel a módszerrel). A rákövetkező években aztán Nirenberg kidolgozott ugyan egy módszert, amely lehetővé tette az egyes

aminosavakhoz tartozó kodonok nukleotidösszetételének megállapítását, de a következő igazi áttörés 1964-ig váratott magára.

A kodonok döntő hányadának jelentését 1964 és 1966 között Nirenberg Philip Lederrel együttműködve fejtette meg. Módszerük a következő volt: ismert szekvenciájú tripleteket, tRNS-t, valamint olyan aminosav-keveréket adtak riboszómákhoz, melyben egy adott aminosav a húszból radioaktívan jelölve volt. A következőekben hagyták, hogy a riboszómákhoz az adott ismert triplet által meghatározott módon egy aminoacil-tRNS kötődjön, majd ezt követően a reakcióelegyet olyan szűrőn vitték át, amely csak a szabad, nem kötött aminoacil-tRNS-eket engedte át. Amennyiben a szűrőn maradó, riboszómák által kötött aminoacil-tRNS-eket tartalmazó frakció radioaktív jelölést mutatott, levonható volt a következtetés, hogy a kezdetben hozzáadott ismert szekvenciájú triplet kódolja az adott aminosavat. A módszerrel a kodonok többségének jelentését sikerült azonosítani, azonban nem az összesnek.

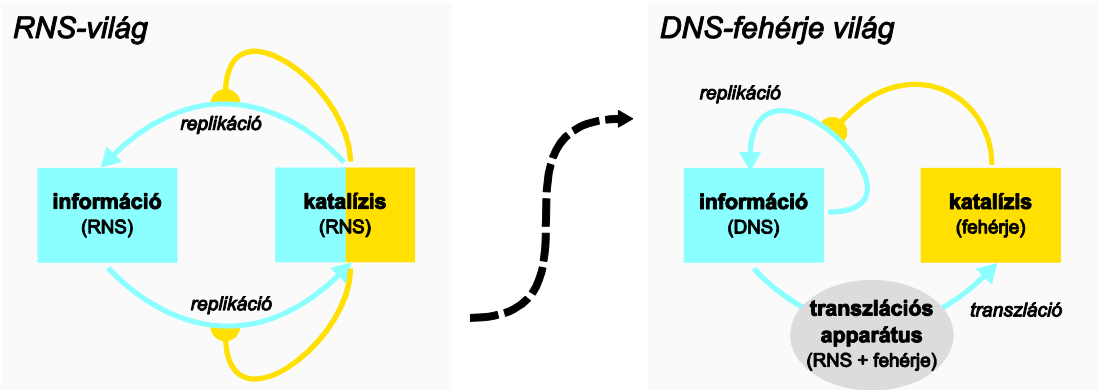
2. ábra: Marshall Nirenberg kísérletezés közben, 1962 körül. A kép forrása: Wikipedia, szabad felhasználású kép.

A teljes kódtábla feltárásának befejező lépései Gobind Khorana munkásságához fűződnek. Khorana olyan szintetikus poliribonukleinsavakat készített, melyekben kettő-, három-, vagy négybázisnyi szekvencia ismétlődött, ezek pedig szabályos ismétlődésű polipeptideket eredményeztek a fehérjeszintézis során. Ezek mintázatát Nirenberg korábbi eredményeivel összevetve, melyekben az egyes kodonok nukleotidösszetételét határozta meg, a még hiányzó kodonok jelentésére is fény derült. 1966-ra így teljessé vált a sztenderd genetikai kód megfejtése.

II. A genetikai kód eredetének problematikája

II.1. A DNS-fehérje világ eredetének kérdése

A ma élő organizmusokat megvizsgálva azt látjuk, hogy két alapvető életműködésüket: a genetikai információ tárolását és továbbörökítését, valamint az anyagcsere-folyamatok lebonyolítását két külön molekulatípus, a DNS és a fehérjék végzik. Mi több, e két molekulatípus léte egymást feltételezi: a megfelelő másolási hűségű DNS-replikáció elengedhetetlen kellékei bizonyos fehérjék – melyek azonban a transzlációs apparátus tevékenysége nyomán a DNS-ben tárolt információ alapján készülnek el. A rendszer tehát körkörös, és amennyiben felvetjük a kérdést, hogy egy ilyen rendszer hogyan jöhetett létre, – Francis Crick szavaival élve – notóriusan nehéz problémához jutunk. Erre a problémára kínál egy lehetséges megoldást az RNS-világ hipotézis, mely azt feltételezi, hogy az élet evolúciójának



3. ábra: Az RNS-világ és a DNS-fehérje világ. Míg az RNS-világban mind az információtárolást, mind a katalitikus tevékenységeket RNS molekulák látják el (a két működés megosztása elképzelhető pl. komplementer szálak között, a kérdés nyitott, vizsgálata jelenleg is folyik), a DNS-fehérje világban ez különválik, és az információ áramlása, mint azt Francis Crick centrális dogmájában megfogalmazta, a katalízis ágensei, vagyis a fehérjék felé egyirányú lesz. A szaggatott nyíl a két világ közti fokozatos, több lépéses, Szathmáry Eörs szerint az evolúció nagy átmenetei közé sorolandó eseménysorra utal. Saját ábra.

korai fázisában, nagyjából 4-3,5 milliárd évvel ezelőtt, létezett egy szakasz, az RNS-világ, amely során az organizmusok központi működéseit, így az információ tárolását és a különböző anyagcsere-folyamatokat egyaránt RNS-molekulák biztosították. Az elmélet kidolgozásának legfőbb empirikus alapjait azok az 1980-as évek elején Sidney Altman és Thomas Cech által elvégzett kísérletek jelentették, amelyek elsőként mutattak ki katalitikus tulajdonságú RNS-molekulákat, vagyis ribozimeket.

A természetes ribozimek létezésének ismeretében, és annak fényében, hogy általánosságban a nukleinsavak igen alkalmas alanyai a templát alapú replikációnak, jogosnak látszik tehát a feltételezés, hogy egy elsősorban nukleinsavakra támaszkodó élet létezhetett, és a jelenlegi DNS-fehérje világ ebből fokozatosan, átmenetek sorozatával alakult ki (3. ábra). Az RNS-világhoz képest a DNS-fehérje világ két fő újítást hoz: egyrészt a DNS, másrészt a transzlációs apparátus, és így a fehérjék létét. A transzlációs apparátus megjelenésének kérdése pedig szükségképpen elvezet a genetikai kód eredetének problémájához is, mely a szakdolgozat fő témája.

A felvázolt, két, lényegileg különböző világot feltételező paradigma elfogadása után azonban a transzláció mechanizmusába ágyazódó genetikai kód eredetének kérdése további problémákat vet fel. A transzlációs apparátus ugyanis egy komplex rendszert jelent, amely egy soklépéses evolúciós eseménysor terméke, és ebben az eseménysorban a genetikai kód kialakulása szükségszerűen egy korai, míg a fehérjék megjelenése minden bizonnyal egy késői történés volt. Ezen gondolatok mentén Szathmáry Eörs a kódoló koenzim fogantyú hipotézist bemutató cikkében mutatott rá (Szathmáry, 1993), hogy a genetikai kód kialakulásának adaptív voltát a fehérjék által kínált, a ribozimekhez viszonyított katalitikus hatékonyságnövekedéssel magyarázni hibás megközelítést jelent. Ez a magyarázat ugyanis figyelmen kívül hagyja, hogy az evolúció „rövidlátó”: ha egy evolúciós változás haszna csak utólag érvényesül, akkor az a változás adaptív folyamatok révén nem tud végbemenni. Szathmáry ezért amellett érvelt, hogy a genetikai kód eredetének tárgyalásakor annak lényegére, mint nukleotid tripletek és aminosavak közti megfeleltetésre koncentrálva a fehérjék helyett a puszta aminosavak által biztosított előnyöket érdemes figyelembe venni, a fehérjék és a transzláció kérdését pedig egy evolúciósan későbbi, külön problémaként kezelni. Fontos megjegyezni, hogy Szathmáry nem állítja, hogy a genetikai kód és a transzláció kérdése ne lenne egy alapjaiban összefüggő kérdéskör, véleménye szerint előbbi az utóbbi számára preadaptációként szolgált (Szathmáry, 1993). Szathmáry szerint továbbá a genetikai kód kialakulása az evolúció nagy átmeneteinek egyike, ugyanis alapjaiban változtatta meg a genetikai információ felhasználásának módját (Szathmáry, 1999). Összefoglalva tehát azt mondhatjuk,

hogy a genetikai kód eredetét a transzláció kialakulásának korai momentumaként, az RNS-világ színfalai közt kell keresnünk.

II.2. Az RNS-világ

II.2.1. Az RNS-világ előtt és után

Ha elfogadjuk, hogy a DNS-fehérje világ előtt létezett az RNS-világ fázisa, akkor logikus kérdésként merül fel, hogy vajon mi előzhette meg magát az RNS-világot. Elképzelhető, hogy az élet eredetét is az RNS-világban kell keresnünk, vagy pedig azt megelőzően az élet egy ismét csak másféle szerveződését kell feltételeznünk? Ezekre a felvetésekre az élet ennyire korai fázisaiból származó nyomok hiányában nem adódnak egyértelmű válaszok (Joyce, 2002). A tudomány jelen állása szerint elképzelhető, hogy az első evolúcióképes – tehát a szaporodás, öröklődés és változékonyság tulajdonságaival bíró – rendszer RNS alapú volt. Mindemellett egyre biztatóbb eredmények születnek arra vonatkozóan, hogy RNS szálak prebiotikus körülmények között létrejöhetnek (Kun, 2011).

A jelenkori szervezetek biokémiáját megvizsgálva a természetes ribozimeken túl is számos olyan jelet találhatunk, melyek egy valamikori RNS-világ létére utalnak. Ezek közül a legszembetűnőbb talán a fehérjeszintézis folyamatában a különböző RNS-félék központi szerepe: a peptidlánc elkészítésének jóformán összes főbb mozzanata, így az aminosavakat összekapcsoló peptidil-transzfer is RNS-ek tevékenysége nyomán megy végbe. A fehérjeszintézis mellett a különböző, alapvető anyagcsere-folyamatokban kulcsfontosságú koenzimek (ATP, NAD, NADP, FAD, koenzim-A) szerkezete is igen szuggesztívnek mutatkozik az RNS-világ kérdésében, ezek ugyanis kivétel nélkül a ribonukleotid adenint tartalmazzák – bár a ribonukleotid rész a koenzimek működésében közvetlenül nem vesz részt (Kun, 2011). Ez utóbbi tény jelentőségére az ősi genetikai kód eredetét tárgyaló kódoló koenzim fogantyú elmélet kapcsán még a szakdolgozat későbbi részében kitérünk. Szintén figyelemreméltó a sejtműködést és azon belül is elsősorban a génexpressziós folyamatokat szabályozó RNS-félék változatossága és központi szerepe. Látható tehát, hogy a különböző RNS-ek a sejtműködések jóformán minden területén

megtalálhatóak, és általában igen beágyazott, alapvető szerepet játszanak azok lefolyásában és szabályozásában.

II.2.2. Az RNS-világ szerveződése

Az RNS-világ élőlényeinek, közkeletű nevükön a ribo-organizmusoknak szerveződése és biokémiájuk potenciális komplexitása a közvetlen bizonyítékok szűkössége miatt fogós, ám mindenképpen lényeges kérdés. A genetikai kód eredetét magyarázó kódoló koenzim fogantyú hipotézis (lásd később) például egy metabolikusan komplex RNS-világ feltételezésére támaszkodik; a genetikai kód eredetének tárgyalását pedig tágabb értelemben is árnyalja az RNS-világ behatóbb ismerete. A fentebb már megemlített, ribonukleotid-résszel rendelkező recens koenzimek léte alapján általánosan elfogadott a következtetés, hogy a ribo-organizmusok a jelenlegi szervezetek biokémiai hálózatának alapjaival már legalábbis rendelkeztek. Több, ennyire közvetlennek tetsző nyom azonban sajnos egyelőre nem áll rendelkezésünkre az egykori RNS-világ metabolizmusának természetét illetőleg, a jelenleg ismert, a kérdésben esetlegesen informatívnak mutatkozó természetes ribozimek ugyanis kivétel nélkül RNS-láncok hasítását végzik (Kun, 2011).

Egy adott tulajdonság, például egy biokémiai útvonal vagy szerveződési sajátosság RNS-világbeli meglétének mérlegeléséhez Gerald F. Joyce (2002) egy komplex szempontrendszert javasol, melynek elsődleges eleme egy adott működés RNS-molekulák általi kivitelezhetőségének vizsgálata. Ennek legkézenfekvőbb, bár kétségtelenül forrásigényes módját jelentik a mesterséges ribozimek evolváltatását célzó kísérletek. A mesterséges ribozimek olyan katalitikus aktivitású RNS-ek, melyeket egy vizsgált jellegre való mesterséges szelekcióval hoznak létre, általában random szekvenciájú RNS-ek populációjából kiindulva. Ezek a kísérletek lehetőséget nyújtanak az RNS-világban, pusztán ribozimek közreműködésével egykor elérhető katalitikus repertoár változatosságának becslésére. A természetes ribozimek felfedezése óta számos mesterséges ribozimet sikerült létrehozni, többek közt olyanokat, amelyek képesek az RNS-molekulák szintézisének lépéseit elvégezni, így ribonukleotidokat összetevőikből előállítani, RNS-láncot aktivált ribonukleotid

hozzákapcsolásával hosszabbítani, valamint a nukleotid aktiválása során elhasznált ATP-t ADP és foszfát felhasználásával regenerálni (Kun, 2011). Problémásnak mutatkozik azonban az RNS-függő RNS-szintézis, vagyis az RNS-replikáció ribozimek általi kivitelezhetősége, ami pedig az RNS-világ egyik központi, bizonyos vélemények szerint (Ellington és mtsai, 2007) pedig egyik legkorábbi működése kellett legyen. Ugyan léteznek RNS-replikációt végző mesterséges ribozimek, ezek azonban legalább két téren is gyengén teljesítenek: az egyik az egy replikációs esemény során lemásolt szakasz hossza, vagyis a processzivitás, a másik pedig a lemásolt információ pontossága, vagyis a másolási hűség (Kun, 2010). A jelenleg ismert legjobban teljesítő ilyen aktivitású ribozim egy templát-primer komplexet 95 darab nukleotiddal, 99%-os másolási hűség mellett egészített ki (Holliger és mtsai, 2011). Lévén, hogy a szóban forgó ribozim 198 nukleotid hosszú, látható, hogy ez a processzivitás nem túl meggyőző. Joyce szerint azonban a ribozimek általi RNS-replikáció egyelőre nyitott kérdése nem igazán súlyos probléma az RNS-világ kérdésében, több okból is (Joyce, 2002). Egyrészt nem kizárt, hogy az RNS-világbeli RNS-replikáció útját nem a biopolimerek másolásának jelenleg általános, monomerek egyenkénti hozzáadásával történő módja jelentette, azonos eredményre vezet például külön-külön elkészített oligonukleotidok templáthoz való asszociációja majd összekapcsolása is. Ez utóbbi aktivitást végző ribozimet sikerrel elő is állítottak, mégpedig egy recens intron-szakasz módosításával (Ellington és mtsai, 2007). Az így kapott ribozim nagyjából 10 nukleotid hosszúságú oligonukleotidokat felhasználva képesnek mutatkozott egy kb. 200 bázispár hosszúságú RNS templát lemásolására (Ellington és mtsai, 2007). Az alternatív replikációs lehetőségek mellett a meggyőző RNS-replikáz ribozimek hiányának problémáját tovább enyhíti Joyce szerint az, hogy már a csupán pár tíz nukleotid hosszúságú ribozimek által bejárható szekvenciatér is elképesztően nagy, a vonatkozó mesterséges szelekciós eljárások pedig ennek minden bizonnyal csak igen kis részét érintették (Joyce, 2002).

Felmerül a kérdés, hogy vajon mutattak-e sejtes szerveződést a ribo-organizmusok. Erre nagy bizonyossággal mostani ismereteink alapján nem lehetséges választ adni, a sejtes szerveződés lehetősége azonban nem kizárt: membránon át történő anyagszállítást végző mesterséges ribozimeket ismerünk (Kun, 2011), a zsírsavak és

így a membránalkotó lipidek szintézisének kulcslépését jelentő Claisen-kondenzáció pedig ribozimek által megoldott (Ellington és mtsai, 2002). Egy további fontos szempontot jelent a különböző energia- és tápanyag-hasznosító folyamatok ribozimek általi katalízisének kérdése. A kísérletes eredmények alapján elmondható, hogy a cukormetabolizmusban központi aldol kondenzációk, a különböző redox folyamatok, valamint változatos (többek között acil-) transzferek a ribozimek nyújtotta lehetőségek keretein belül vannak (Ellington és mtsai, 2007). Ezek mellett ismertek ribozimek, melyek transzmetilációkat és transzészterifikációkat katalizálnak (Szathmáry, 1993). Mindezen megfontolások alapján levonható tehát a konklúzió, hogy az egykori ribo-organizmusok figyelemreméltó komplexitást érhettek el ribozimeik révén.

III. Az ősi genetikai kód kialakulása

III.1. A kódoló koenzim fogantyú hipotézis

A kódoló koenzim fogantyú hipotézis szerint a genetikai kód eredetét egy hangsúlyosan komplex metabolizmusú RNS-világ színfalai között kell keresnünk (Szathmáry, 1999). Egy ilyen, bonyolult és változatos biokémiai folyamatokkal bíró stádium, mint az a megelőző fejezetben láttuk, jogos felvetésnek tekinthető. Egy további, már az előző részben is említett utalást a hajdani RNS-világ kémiai komplexitására a jelenlegi DNS-fehérje világ adenint, vagyis ribonukleotidot tartalmazó koenzimei jelentik. Ezek ugyanis változatos csoportjaik révén igen alapvető és összetett, tehát komplex biokémiát feltételező metabolikus folyamatok elengedhetetlen szereplői, ribonukleotid részük viszont érdekes módon ezekben a folyamatokban nem vesz részt. Felmerül a kérdés tehát, hogy mi lehet, vagy éppenséggel mi lehetett a nukleotid rész funkciója ezekben a molekulákban. A válasz Szathmáry szerint éppen egy komplex RNS-világban rejlik: a ribonukleotid rész, érvelése szerint, egy fogantyú (vagy jelenleg inkább annak vesztigiális maradéka), melyen keresztül az egykori ribo-organizmusok ribozimei az egyes koenzimeket bázispárosodások révén manipulálni tudták. Ez a kapcsolat az egyes ribo-organizmusok számára egy komplex kémiájú környezetben adaptív előnyt biztosított, hiszen így olyan csoportok váltak részükre elérhetővé, melyek nagyban elősegíthették az egyes biokémiai folyamatok katalízisét. Tanulságos, ha megvizsgáljuk, hogy a ribozimek önmagukban milyen kevés féle csoporttal, kémiai reakciók szempontjából érdekes jelleggel bírnak: képesek hidrogén-hidat képezni, vannak hidrofób csoportjaik és van cukor részük (Szathmáry, 1993). Kétségkívül előnyösnek tűnik tehát a nagyobb változatosságot biztosító csoportok hasznosításának kialakulása. Egy további szempont, hogy a koenzimeket hasznosító ribozimek léte kísérletesen is alátámasztott (Szathmáry, 1999).

III.1.2. A nukleinsav-aminosav hozzárendelés megjelenése

Mint említettük, az RNS-világ ribozimeinek katalitikus repertoárját nagyban szélesíthette a különböző koenzimek felhasználása. Szathmáry javaslata szerint az

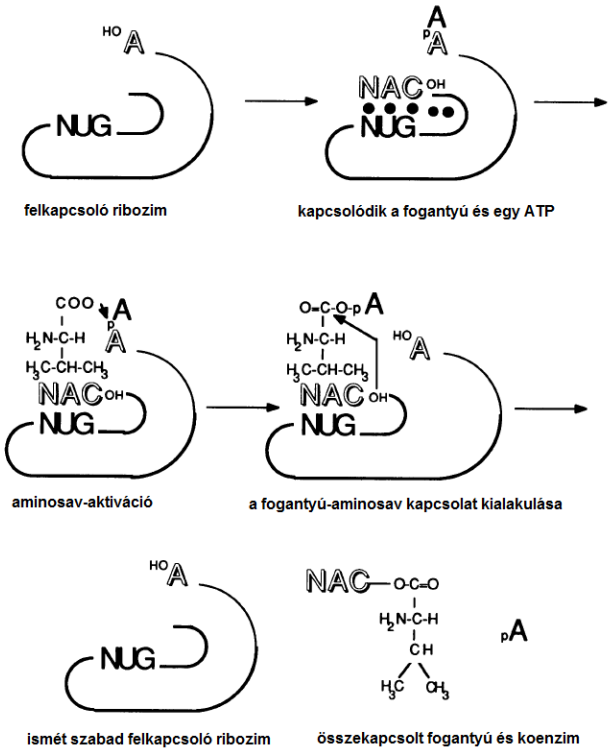
aminosavak hasznosítása is ilyen módon kezdődött. Az aminosavak által biztosított kémiai változatosság, főként a puszta RNS molekulákkal összevetve, jelentős: egyebek mellett alifás amint, karboxilt, imidazolt és tiolt hordoznak (Szathmáry, 1993). Ezen csoportok hasznosítása ribozimek által két féle módon képzelhető el. Az egyik esetben azokon a ribozimeken, melyek aktivitását egy adott csoport elősegíti, erre a csoportra specifikus felismerő-kötőhely alakul ki. Ez azonban több szempontból is problémásnak, és így evolúciós szempontból előnytelennek bizonyulhat. Egyrészt minden érintett ribozimen létre kell jönnie egy specifikus kötőhelynek, másrészt így az adott aminosav esetleges adaptív megváltozása (például szintézisútvonalának kiegészülése révén) igen nehézzé válik. Ez ugyanis azt igényelné, hogy az összes olyan ribozim kötőhelye, amely az adott aminosavra jellemző csoportot használja, egyszerre, ugyanúgy változzon ezzel párhuzamosan, hogy a megváltozott aminosav felismerése és felhasználása továbbra is biztosított legyen. Egy ilyen sokkomponensű, egyszerre végbemenő, egy irányba mutató evolúciós változás nagyon valószínűtlen (Szathmáry, 1999). Erre a problémára kínálnak megoldást az aminosavakat hordozó ribonukleotid kódoló fogantyúk, melyek az aminosavak koenzimként való hasznosításának másik útját kínálják. A ribozimek az elmélet által javasolt kódoló fogantyúkat konvencionális Watson-Crick bázispárosodással ismerik fel és kötik, ez pedig lehetőséget kínál a fogantyúkon aminosav-specifikus szekvenciák kialakulására. Ezen szekvenciák kulcsszerepe, hogy lehetővé teszik az aminosav koenzimek ribozimek általi egyszerű, formális azonosítását, és így moduláris, megismételhető, és megbízható felhasználását. A kódoló fogantyúk révén ez a genetikai kód, mint nukleinsavak és aminosavak közti egyértelmű hozzárendelés evolúciós megjelenésének momentuma (Szathmáry, 1999). Az aminosavak felismerésének problémája természetesen nem tűnik el: ennek feladata specifikus kapcsoló ribozimekre hárul, melyek az egyes aminosavakat a megfelelő szekvenciájú fogantyúval kovalensen összekötik. Ez azonban aminosavanként csak egy féle ribozim egyetlen felismerő-kötőhelyének kialakulását igényli, amely az aminosav esetleges evolúciós megváltozásakor könnyedén alkalmazkodhat. Szathmáry felveti továbbá, hogy már a fogantyúk és aminosavak között létrejövő hozzárendelésre jellemző lehetett a genetikai kód egyik fő tulajdonsága: az egyértelműséggel kombinált degeneráltság, vagy redundancia

(Szathmáry, 1999). Ez a kódoló fogantyú elmélet keretei között azt jelenti, hogy egy aminosavhoz több fogantyú is tartozik, egy fogantyú viszont csak egy féle aminosavat hordozhat. Ennek adaptív volta kettős: egy adott aminosav így szélesebb körben lesz felhasználható (hiszen a több különböző fogantyú több lehetőséget biztosít a hasznosításra), viszont egy adott fogantyú megbízhatóan csak egy féle aminosavat kapcsol egy adott ribozimhez (ha nem így lenne, az a katalitikus hatékonyság romlását eredményezhetné).

Az elmélet talán legkevésbé tisztázott része a kódoló koenzim fogantyúk elsődleges megjelenésének oka és mikéntje. A ribozimek által hasznosított kódoló fogantyúk jelensége, hasonlóan a fehérjeszintézishez, egy több komponensű, összetett rendszert alkot, melynek létrejöttéhez mindenképpen szükségesnek tűnik valamiféle preadaptációs előzmény. Szathmáry arra utal, hogy ez esetleg az aminosavakat létrehozó szintézisutak optimalizációja során alakulhatott ki, amikor is a köztes metabolitokat a különböző ribozimek egy ribonukleotid fogantyú révén manipulálhatták (Szathmáry, 1999). A feltevést erősíti, hogy a már kódoló fogantyúk és az egyes szintézisutak között egyértelmű kapcsolat látszik (Szathmáry, 1994). Erről több szó a kód kiterjesztését és optimalizációját tárgyaló fejezetekben fog esni.

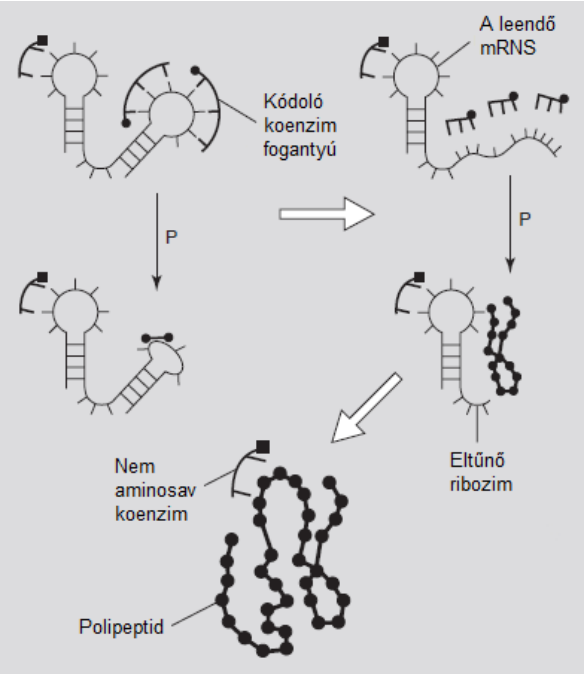
A kódoló koenzim fogantyúhoz Szathmáry elképzelései szerint egy felkapcsoló ribozim tevékenysége nyomán kapcsolódik egy adott aminosav. A kapcsolatot az elmélet szerint egy kovalens, nitrogénen és karbamoil csoporton keresztül megvalósuló kötés jelentette (Szathmáry és mtsai, 2007). Fontos kérdést jelent, hogy maguk a felkapcsoló ribozimek milyen módon azonosították és kapcsolták fel az egyes aminosavakat. Szathmáry szerint ehhez először a fogantyúknak kell specifikus bázispárosodás révén kötődnie a megfelelő felkapcsoló ribozimhez. Ezt követően a megfelelő aminosav felismerését a ribozim egyik, úgynevezett diszkriminátor bázisa és a fogantyú specifikus szekvenciája együttesen, sztereokémiai alapon végzik el. Amennyiben a felismerés megtörtént, a felkapcsoló ribozim az aminosavat egy ATP elhasználásával felaktiválja, majd összekapcsolja a fogantyúval. Végül az elkészült fogantyú-aminosav kettős, valamint egy AMP távozik (Szathmáry, 1993; 4. ábra).

A kódoló koenzim fogantyú az elmélet szerint egy hajtű szerkezetű oligoribonukleotid volt, melyet az egyes ribozimek Watson-Crick bázispárosodással kötöttek és ismertek fel. A mai kódot figyelembe véve ez a felismerés talán tripletes volt, bár ez kérdéses (Szathmáry, 1999). Az viszont kísérletes eredmények alapján biztosnak látszik, hogy egy, a ribozim katalitikus funkciójának elősegítéséhez szükséges stabilitású kapcsolatot pusztán tripletes kötődés nem tesz lehetővé. A bázispárosodást tehát kiegészítendő, elképzelhető, hogy a ribozimek a kódoló fogantyúkkal egy evolvált kötőhely révén teremtettek kapcsolatot, melyet feltehetően a bázisok egymásra lapolódásából eredő (angol szakkifejezéssel *base-stacking*) kölcsönhatások stabilizáltak. Ezt szintén alátámasztják kísérleti eredmények (Szathmáry, 1999). Az evolvált kötőhely egy további előnye, hogy kizárja egy-egy kódoló fogantyú véletlen bekötődését valamely ribozimre, csak azért, mert azon a megfelelő, tehát a kódoló fogantyú szekvenciájával komplementer szekvencia megtalálható (Szathmáry, 1999). A fogantyú hajtű szerkezetéről továbbá megemlítendő, hogy ez a túl korai lebomlás gátja lehet, amely a hatékony működés szempontjából mindenképp előnyös (Szathmáry, 1999).



4. ábra: A kódoló fogantyú és az aminosav kapcsolatának kialakulása. A részleteket lásd a szövegben. A kép forrása: Szathmáry, 1993.

Mindenképpen szükséges kitérnünk a nem aminosav koenzimek kérdéskörére, már csak azért is, mert többek között ezek vesztigiálisnak ható ribonukleotid részei adták a kódoló koenzim fogantyú elmélet kidolgozásának ötletét (Szathmáry, 1999). Szathmáry szerint ezek azért nem tudtak a genetikai kód és az aminosavak által



5. ábra: Egy lehetséges út a kódoló koenzim fogantyúktól a polipeptidekig. A fekete nyilak biokémiai modifikációkat (peptidil transzfert), a fehér nyilak pedig evolúciós átmeneteket jelölnek. A kép forrása: Szathmáry, 1999., módosítva.

bejárt, végül a transzlációs apparátushoz és fehérjékhez vezető úton már csak elindulni sem, mert az aminosavakkal ellentétben nem tartoznak egy reguláris molekulaosztályba, így egy általános molekuláris apparátus esetükben nem tudott kialakulni (Szathmáry, 1999).

Noha az elmélet hangsúlyosan nem érinti a transzláció problémáját (Szathmáry, 1993), keretei mégis lehetőséget nyújtanak a majdani transzlációs apparátus főbb elemeinek azonosítására. Eszerint a kódoló fogantyú a tRNS előfutára, specifikus szekvenciája pedig az antikodonnak feleltethető meg. Az egyes ribozimek, melyek evolvált kötőhelyeiken kodonokat viselnek, átalakulással az mRNS-ek elődjeinek tekinthetők, katalitikus funkciójukat pedig a hozzájuk asszociálódó aminosavak összekapcsolódásával egy RNS-fehérje világ átmeneti stádiumaiban egyre komplexebbé váló polipeptidek vehették át (lásd 5. ábra). Többek között ezen az úton jöhettek létre a felkapcsoló ribozimekből a mai aminoacil-tRNS-szintetázok is.

III.2. A sztereokémiai hipotézis

A sztereokémiai hipotézis szerint a genetikai kód kialakulása során létezett egy „sztereokémiai éra”, melynek során a nukleinsavak és az aminosavak közti első asszociációk megjelenésének hátterében elsősorban az ezek közt specializált kötőhelyeken kialakuló sztereokémiai kölcsönhatások álltak (Yarus és mtsai, 2009). Az elmélet legfőbb predikciója, hogy ha ennek az időszaknak maradtak nyomai a jelenlegi, sztenderd kódban, azok kimutathatóak, mégpedig mesterséges aminosav-kötő ribozimek evolváltatásával. Amennyiben ugyanis egy adott aminosav kötésére

adaptálódott RNS-kötőhely szekvenciájában az aminosav valamely kodonja vagy antikodonja a véletlen szintnél nagyobb arányban fordul elő, akkor adódik a következtetés, hogy ez az asszociáció a korai aminosav-kötő ribozimek specifikus kötőhelyeinek szekvenciájából származik (Yarus és mtsai, 2009).

Ezen meggondolások alapján foglalták össze Michael Yarus és munkatársai 2009-ben a vonatkozó kísérletes eredményeket, különböző evolvált aminosav-kötő RNS struktúrák szekvenciáinak összegyűjtésével és elemzésével (Yarus és mtsai, 2009). Statisztikai módszerekkel azt találták, hogy a feltételezett sztereokémiai éra nyomai, elsősorban az antikodonok kötőhelyekben való kiemelkedő gyakoriságának formájában bizonyos aminosavak, például hisztidin vagy triptofán esetén kimutathatóak. Eredményeik mindazonáltal ellentmondásosak, ugyanis az elemzés csak nyolc aminosav esetét vizsgálta, így a jelenség valós súlyáról és jelentőségéről keveset árul el.

III.3. A genetikai kód kiterjeszkedése

A jelenleg ismert sztenderd genetikai kód kialakulásának igen fontos fejezetét jelentette a kód kiterjeszkedése, aminek során a genetikai kód által kódolt aminosavak száma a kezdeti néhányról a jelenlegi húszra emelkedett. Mint azt a pirrolizin és a szelenocisztein aminosavak bizonyos szerveztekben való kódolása, valamint a kódolt aminosav-repertoár szintetikus kiterjesztésének relatíve könnyű volta is mutatja, ez a folyamat feltehetően nem állt le véglegesen, bár intenzitása az új aminosavak belépése által jelentett átmeneti rátermettség-csökkenés miatt erősen csökkent (Wong, 2005).

A kód kiterjeszkedésének kérdését árnyalja, hogy arra csupán a kódolt aminosavak repertoárjának bővüléseként, vagy ezzel párhuzamosan a kódoló funkcióval bíró kodonok számának növekedéseként is tekintünk. Ebben a kérdésben a vélemények megoszlanak. Bizonyos nézetek szerint már a korai kód is felhasználta mind a 64 darab lehetséges kodont, értelemszerűen a mai, sztenderd kódra jellemzőnél nagyobb fokú redundancia mellett (Koonin és mtsai, 2009). Mások szerint az ősi genetikai kódban a funkcionális kodonok száma is kevesebb volt a mainál, és a kód

kiterjeszkedésének a kódoló funkciójú kodonok párosával történő számbeli növekedése is része, mely folyamatnak az elsődleges meghatározója az egyes kodonpárok csökkenő termodinamikai stabilitása volt (Trifonov, 2004). A két megközelítést elsősorban az differenciálja, hogy feltételeznek-e működő transzlációt a korai kód evolúciójának idején, hiszen egy olyan kód esetén, amelyben bizonyos kodonok nem értelmezhetőek, egyéb *ad hoc* körülmények feltételezése nélkül a transzláció problémás lehet. Eugene V. Koonin és munkatársai (Koonin és mtsai, 2009) közleményükben egy már meglévő, de még kezdetleges, kis pontosságú transzlációs apparátus mellett foglalnak állást. Trifonov ezzel szemben nem tér ki a transzláció kérdésére (Trifonov, 2004), bár álláspontja az előbb mondottak szerint implikálja, hogy a korai kód evolúciója a transzláció létrejöttét megelőzően ment végbe. A téma további tárgyalása során Koonin és munkatársai nézetének helyességét tesszük fel, és a kód kiterjeszkedése alatt kizárólag a proteinogén aminosavak repertoárjának növekedését értjük.

III.3.1. A kód kiterjeszkedésének adaptív háttere

Mint azt a kódoló koenzim fogantyú elmélet tárgyalásakor is hangsúlyoztuk, mindenképpen szükséges a különböző evolúciós folyamatokat kísérő, azonnal érvényesülő adaptív jellegek azonosítása. A genetikai kód kiterjeszkedése során, némiképp a kódoló koenzim fogantyú elméletből kiindulva egyrészt feltételezhetjük, hogy az új aminosavak pusztán koenzimként funkcionálva kínáltak előnyös kiegészítéseket egy adott szervezet biokémiája számára. Másrészt viszont abból kiindulva, hogy a kezdeti néhány kodon-aminosav asszociáció kialakulását követően feltehetően immáron a korai fehérjék, vagy legalábbis polipeptidek szintézisét végző apparátus kialakulása is megkezdődhetett, az új aminosavak befoglalása a kódba ezen polipeptidek adaptivitását is növelhette. A szekvencia megváltozásával az adaptivitás növekedése polipeptidek, illetve fehérjék esetén két fő hatás révén történhet: az egyik a térbeli struktúra stabilitásán, a másik a katalitikus aktivitás finomhangolásán keresztül érvényesül (Müller és mtsai, 2013). Amennyiben feltételezzük, hogy az új aminosavak használata elsődlegesen a fehérjék adaptivitásának növekedésén keresztül bizonyult előnyösnek, úgy ennek folyományaként azt is szükséges feltennünk, hogy egy, a jelenlegihez képest erősen

korlátozott aminosav-készlet felhasználásával is lehetséges funkcionális fehérjéket előállítani. Erre egyrészt élővilágbeli példák, másrészt kísérletes bizonyítékok vannak. A lepényhalfélék családjába tartozó *Pseudopleuronectes americanus* szervezetében található, fagyhatások kivédését szolgáló fehérje például csupán hét különböző aminosavból épül fel; egy kísérlet során pedig egy mesterségesen előállított, mindössze kilenc különböző aminosavból álló korizminsav mutáz enzim a módosítás ellenére megtartotta aktivitását, bár hatékonysága és stabilitása érzékelhetően csökkent (Müller és mtsai, 2013).

A kód kiterjeszkedését minden valószínűség szerint az új aminosavak által kínált előnyökkel párhuzamosan egy sor adaptív hatás, illetve környezeti körülményekből adódó szükségszerűség befolyásolta. Az előbbiek közül a kód transzlációs hibákkal szembeni optimalizálódásának kérdésére a szakdolgozat negyedik fejezetében térünk ki, míg az utóbbiak közül az aminosavak elérhetőségéről és egyéb folyamatokról a következőekben szólunk.

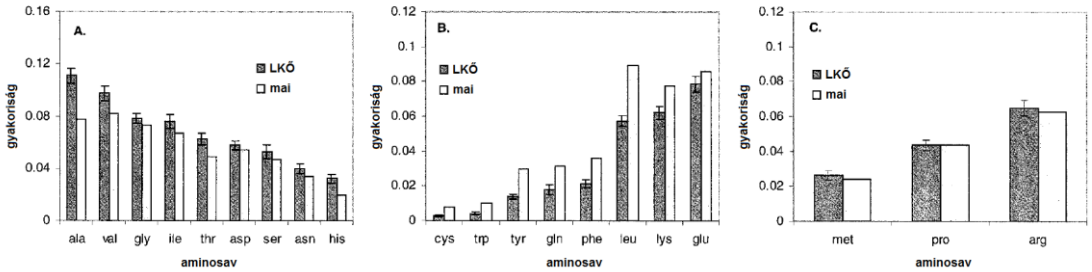
III.3.2. A korai aminosavak halmazának azonosítása

A genetikai kód kiterjeszkedésének tárgyalásakor logikus kiindulópontként kínálkozik a kód megjelenésének időszakában elérhető és a kódba így feltehetően korán bekerülő aminosavak halmazának azonosítása. Noha a vonatkozó kutatások általában az aminosavak belépésének egzakt sorrendét is megkísérlik rekonstruálni, mi a témában mutatkozó konzenszus hiányában ettől inkább eltekintünk.

A korai, avagy prebiotikus aminosavak kérdésének legelső kísérletes megközelítését Stanley Miller és Harold Urey 1952-végzett kísérleteinek 1953-ban közölt, azóta pedig klasszikussá vált eredményei jelentik (Miller, 1953). Az eredeti kísérlet során egy körkörös áramlású, légürített apparátusba vizet helyeztek, ehhez pedig a korai Föld atmoszférájáról szóló korabeli elképzelésekkel összhangban metán, ammónia és hidrogéngáz keverékét adták. Ezt követően az edényt lezárták, a vizet forralni kezdték, az így keringeni kezdő gázelegyet pedig, az atmoszferikus villámlást imitálandó, elektromos kisüléseknek tették ki. A gázelegyet végül minden egyes körben lecsapatták, a kondenzátumból pedig azt vízen átvezetve szűrték ki a

várakozások szerint létrejövő komplex, és így a kísérlet hőmérsékletén el nem párolgó molekulákat. A procedúra végeztével az edényből mintát vettek és annak összetételét kromatográfiás eljárással elemezték. Noha a legelső vizsgálatok a proteinogén aminosavak közül csak aszparaginsav, glicin és alanin jelenlétét mutatták ki (Miller, 1953), a rákövetkező évek-évtizedek kísérletei összesen tíz proteinogén aminosav prebiotikus szintézisének a lehetőségét mutatták meg. Ezek: glicin, alanin, aszparaginsav, glutaminsav, valin, szerin, izoleucin, leucin, prolin, és treonin (Koonin és mtsai, 2009).

A kísérletes és elméleti megközelítések határmezsgyéjén mozog Dawn J. Brooks és munkatársainak elemzése, melyben statisztikai módszerek alkalmazásával kíséreltek meg becslést tenni a legutolsó közös ős fehérjéinek aminosav-összetételére (Brooks és mtsai, 2002). Ennek során a legutolsó közös ős feltételezett fehérjekészletéből, melyet korábbi munkák összehasonlító genomi elemzések alapján állítottak össze, 65 darab fehérjét választottak ki, majd ezek ősinek vélt szekvenciáit az egyes alkotó aminosavak konzerválódási valószínűségének fényében, számítógépes szimulációk futtatásával vizsgálták. Az eljárással három aminosav-csoportot különítettek el: egy olyat, amely tagjainak a recens fehérjékben való aránya az ősi összetételhez képest növekedett, egyet, amely tagjainak aránya csökkent, és egyet, amely tagjainál nem volt kimutatható változás az ősi és a mai állapot között (lásd 6. ábra). Ha feltesszük, hogy azon aminosavak aránya volt nagyobb a legutolsó közös ős fehérjéiben, amelyek számára az adott időszakban elérhetőek voltak, akkor a korai aminosavak egy újabb feltételezett listájához jutunk. Ez: alanin, valin, glicin, izoleucin, treonin, aszparaginsav, szerin, aszparagin, és hisztidin (Brooks és mtsai, 2002).



6. ábra: Aminosavgyakoriságok a legutolsó közös ős (LKŐ) és a mai organizmusok fehérjéiben. Az A. rész azon aminosavakat mutatja, melyek gyakorisága csökkent, a B. rész aminosavainak gyakorisága nőtt, a C. részben szereplő aminosavak gyakorisága pedig nem változott. A kép forrása: Brooks és mtsai, 2002.

Módszerében talán vitatható, mindazonáltal érdekes eredményre vezet Edward N. Trifonov elemzése. Trifonov összesen hatvan, az aminosavak genetikai kódba való belépésének sorrendjét meghatározni kívánó munka és egyéb vonatkozó kutatás eredményét egységesítette egy konszenzus aminosav-sorrend felállítása érdekében (Trifonov, 2004). Eredményei, nem meglepő módon, jórészt összecsengenek a korai aminosav-készletre vonatkozó eddigi bemutatott becslésekkel. A Trifonov által megalkotott konszenzus-sorrend első tíz tagja: glicin, alanin, aszparaginsav, valin, prolin, szerin, glutaminsav, treonin, leucin, arginin (Trifonov, 2004).

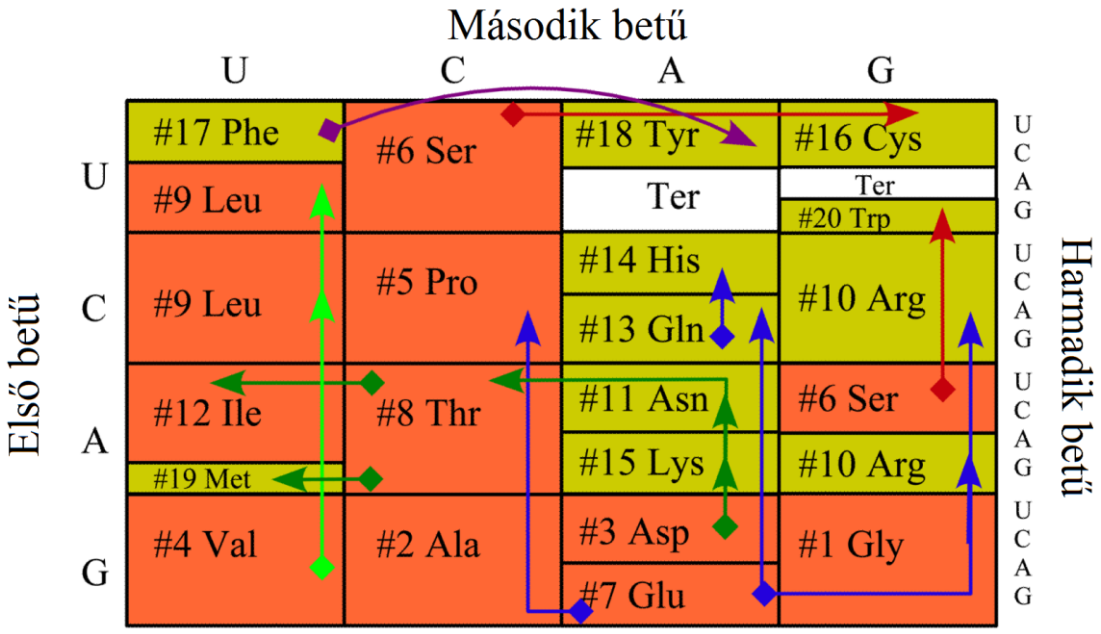
aminosav	*Miller-Urey*	*Brooks*	*Trifonov*	*KEH*
Gly	+	+	+	+
Ala	+	+	+	+
Val	+	+	+	+
Ile	+	+	–	+
Leu	+	–	+	+
Thr	+	+	+	+
Ser	+	+	+	+
Asp	+	+	+	+
Asn	–	+	–	–
Glu	+	–	+	+
His	–	+	–	–
Pro	+	–	+	+
Arg	–	–	+	–

1. táblázat: A korai aminosavak halmazára vonatkozó három megközelítés, valamint a koevolúciós hipotézis (KEH, lásd lejjebb) által elsődlegesnek tartott aminosavak listájának összehasonlítása. Az adatok forrását lásd a szövegben.

III.3.3. A koevolúciós hipotézis

A J. Tzei-Fei Wong által 1975-ben bemutatott koevolúciós hipotézis központi állítása, hogy lévén a proteinogén aminosavak teljes készlete nem lehet prebiotikus eredetű, ezért az aminosav-repertoár kibővülését párhuzamosan a releváns bioszintetikus útvonalak evolúciója kísérte. Ez pedig az elmélet proponensei szerint a

genetikai kód evolúciójával (értve ez alatt az adott kodon-aminosav asszociációk megváltozását) szoros összefüggésben történt, kölcsönös egymásra hatások mellett (Wong, 2005). Egyszóval tehát az elmélet szerint a kód kiterjeszkedése során a genetikai kód és az aminosavak egy részét létrehozó bioszintetikus útvonalak koevolúciója ment végbe.



7. ábra: A genetikai kód kiterjeszkedése a koevolúciós elmélet szerint. A 1. fázis aminosavai narancssárgával, a 2. fázis aminosavai pedig zölddel vannak jelölve. Az egyes aminosavak közti bioszintetikus kapcsolatokat a nyilak jelölik, a különböző szín pedig különböző bioszintetikus családokra utal (glutaminsav – kék, aszparaginsav – sötétzöld, fenilalanin – magenta, szerin – piros, és valin – világoszöld). A kép forrása: Koonin és Novozhilov, 2009.

Az elmélet kidolgozása során a jelenlegi bioszintetikus útvonalak tanulmányozása alapján prekurzor-termék kapcsolatokat állapítottak meg az egyes aminosavak között, ezek összegzésével pedig felállították az aminosavak feltételezhető genealógiáját (lásd 7. ábra). A koevolúciós hipotézis az élő szervezetek által valamilyen módon felhasznált aminosavakat három kategóriába sorolja. Az 1. fázis aminosavai közé azok tartoznak, amelyek elsődlegesen a prebiotikus szintézis termékeiként kerültek be a kódba (Wong, 2005). Igen látványos módon ez a csoport, melynek összetétele kizárólag a metabolikus útvonalak elemzésén keresztül lett meghatározva, majdnem tökéletesen egybeesik az egyéb kutatások, például prebiotikus szintézisek imitációja alapján elsődlegesnek tartott aminosavak körével (Wong, 2005; Koonin és Novozhilov, 2009). A koevolúciós hipotézis 1. fázisának

aminosavai: glicin, alanin, aszparaginsav, valin, prolin, szerin, glutaminsav, treonin, leucin, és izoleucin (Koonin és Novozhilov, 2009). A 2. fázis aminosavait azok képezik, amelyek az 1. fázis tagjainak bioszintetikus leszármazottai (Wong, 2005). Ezek: arginin, aszparagin, glutamin, hisztidin, lizin, cisztein, fenilalanin, tirozin, metionin, és triptofán (Koonin és Novozhilov, 2009). A 3. fázis aminosavai pedig azok, amelyek a sztenderd genetikai kód által nem kódoltak, fehérjékben való jelenlétük így poszt-transzlációs modifikációknak köszönhető. Ilyen például a hidroxiprolin, az adenililtirozin, a pirroglutaminsav, vagy a glikozil-aszparagin (Wong, 2005). Az 1. és a 2. fázis aminosavairól, valamint a köztük fennálló bioszintetikus kapcsolatokról lásd a 7. ábrát.

A koevolúciós hipotézis az aminosav-repertoár bővülését az ún. inventív bioszintézis jelenségén keresztül képzeli el (Wong, 2005). Az inventív bioszintézis a bioszintetikus útvonalak evolúciójának azt a folyamatát jelöli, amely során ezen szintézisutak kiterjedése egy organizmus rátermettségét különböző, addig nem használt biokémiai vegyületek megjelenése révén növeli. Ennek a jelenségnek egy, aminosavak és származékaik esetére vonatkoztatható alesetét nevezi a koevolúciós hipotézis pretran-szintézisnek (Wong, 2005). A pretran-szintézis neve a pretranszlációs szintézis összevonásából jön, és ennek megfelelően azt értjük alatta, amikor egy prekurzor aminosav a neki megfelelő tRNS-hez kötve, tehát aminoacil-tRNS formában, a transzlációt megelőzően szolgál egy biokémiai átalakítás szubsztrátjaként. A reakció eredménye egy termék aminoacil-tRNS, amely aztán a riboszómára az antikodon változatlansága miatt a prekurzor aminosavnak megfelelő kodon beolvasásakor kerül, és épül be ezt követően egy készülő fehérjébe. A pretran-szintézis folyamatára recens példák is vannak, így épül be például fehérjékbe a szerin-tRNS átalakulása után a szelenocisztein, illetve az amidotranszferáz enzim tevékenysége nyomán így készül bizonyos esetekben az Archaea és Bacteria birodalomba tartozó szervezetekben a glutaminsav-tRNS-ből glutamin-tRNS, az aszparaginsav-tRNS-ből pedig aszparagin-tRNS. Tovább árnyalja a pretran-szintézis szerepét, hogy az így létrehozott aminoacil-tRNS származékok egy része aztán nem is a transzláció, hanem egyéb folyamatok, például a porfirin-szintézis vagy sejtfali peptidek szintézise révén hasznosul (Wong, 2005).

A koevolúciós hipotézis feltevése szerint a pretran-szintézis egyfajta előkészítő, megelőző szerepet játszott a 2. fázisba tartozó aminosavak kódba való belépésekor. Egy aminosav belépésének kezdetén a kialakuló pretran-szintézis az antikodon változatlansága miatt a transzláció bizonyos fokú bizonytalanságát eredményezte. Ezen a ponton két, az adaptivitást ellentétesen befolyásoló tényezőt azonosíthatunk: egyrészt az transzláció bizonytalanságából következően a prekurzor aminosavat felhasználó fehérjék stabilitásának és katalitikus hatékonyságának csökkenését, másrészt a termék aminosav által nyújtott előnyöket. Ahhoz, hogy egy aminosav a taglalt úton sikerrel léphessen be a kódba, értelemszerűen az utóbbi hatás dominanciája volt szükséges. Amennyiben tehát az adaptivitást meghatározó tényezők kedvezően alakultak, megkezdődött az új, termék aminosav fokozatos integrációja a genetikai kódba és az adott (és az esetek döntő többségében a legutolsó közös ősnél korábbi) leszármazási vonal biokémiájába. Az átmeneti fázis lezárultát végül a prekurzor aminosav tRNS-ének és aminoacil-tRNS-szintetázának duplikációja, majd ezekből a termék aminosavhoz tartozó paralógok evolúciója jelentette. Ilyen paralógiára mutatnak példát Wong szerint az aszparagin-tRNS szintetáz és az aszparaginsav-tRNS szintetáz enzimek (Wong, 2005).

Noha koevolúciós hipotézist általánosságban a genetikai kód kialakulását magyarázó legfontosabb elméletek közt emlegetik (Szathmáry, 1999; Freeland és mtsai, 2003; Koonin és Novozhilov, 2009), megítélése közel sem egységes. Az elmélet kritikusai szerint bár valószínűsíthető, hogy a kód kialakulását a koevolúciós hipotézis által felvázolt folyamatok jelentősen és alapvetően befolyásolták, mégis korai Wong azon megállapítása, hogy „a koevolúciós hipotézis egy bizonyított tény”; továbbá felhívják a figyelmet az aminosavak közt megállapított prekurzor-termék kapcsolatok szubjektív, nem kellően megalapozott voltára (Koonin és Novozhilov, 2009). Wong ezzel szemben az elmélet elsöprő dominanciája mellett tesz hitet: véleménye szerint a koevolúciós folyamatok a genetikai kód alakításában a többi (később részletezett) elmélet által bemutatott hatásokhoz képest nagyjából öt-hét nagyságrenddel nagyobb szerepet játszott (Wong, 2005). Ennek a számszerű megállapításnak a hátterében egy nehezen komolyan vehető számítás áll, amelyre itt

nem térünk ki. Wong azonban egy másik, jóval meggyőzőbb számítást is felhoz a koevolúciós elmélet mellett, amely a pretran-szintézis azon könnyen belátható következményéből indul ki, hogy egy prekurzor aminosav leszármazottai a kódba való integrációt követően a kódtáblában egymáshoz közel helyezkednek el. Ez alapján a koevolúció lenyomatának meglétét a kód struktúráján a következő számítással ellenőrizhetjük: vegyük a 20 + 2 proteinogén aminosav lehetséges összes, 22 x 21 / 2 = 231 darab párosítását, majd nézzük meg, hogy e párok hányad részét adják az elmélet által azonosított módon azonos prekurzortól származó termék („testvér”) aminosavak. Wong szerint 10 ilyen pár van, ami 4,3%-os aránynak felel meg (Wong, 2005). Vegyük ekkor az azonos, négykodonos blokkokban lévő aminosavak lehetséges párosításait, melyekből 10 van, és nézzük meg ebben a testvér aminosavak arányát. Itt Wong 5 párt azonosít, ami 50%-os arányt jelent (Wong, 2005). A két értéket (4,3% és 50%) összevetve látható, hogy az azonos blokkokban elhelyezkedő aminosavak jóval nagyobb arányban származnak közös prekurzortól, ami a koevolúciós hipotézis érvényessége mellett egy erős érvnek tetszik.

IV. A sztenderd genetikai kód mintázatai és a genetikai kód evolúciója

IV.1. A sztenderd genetikai kód mintázatai

A genetikai kód eredetét illetőleg a kialakulás és a kiterjeszkedés mellett egy további nagy kérdéskört jelent a sztenderd, vagyis az élővilágban közel univerzális genetikai kód stuktúrájának kutatása. A vonatkozó vizsgálatok központi kérdése, hogy miért pont ez a kodonkiosztás alakult ki a sztenderd kód evolúciója során. A kérdést kissé kibontva: mik voltak azok a háttérkörülmények, történeti események, és evolúciós tényezők, melyek összjátéka az egyes aminosavaknak és kodonoknak éppen a sztenderd kódtáblában látható asszociáltságát eredményezte, és vajon mi volt e hatások szerepe és viszonya nem csupán az egyes kodon-aminosav kapcsolatokat, hanem az egész kódtáblát tekintve? A lehetséges kódok száma csillagászati (10^{83} és 10^{84} között van), a sztenderd genetikai kód struktúrája pedig, mint azt a már ránézésre is látható blokk-struktúra mutatja (1. ábra), nem véletlenszerű, ezek a kérdések tehát mindenképpen jogosak.

A blokkos struktúra, vagyis a szinonim kodonok kettes, négyes, illetve két esetben hatos csoportokban való elhelyezkedésén túl a sztenderd genetikai kódban számos egyéb mintázat azonosítható. A kód eddig azonosított, további magyarázatot érdemlő mintázatai (Koonin és Novozhilov, 2009): 1. azon kodonok, melyek második nukleotidja U, hidrofób aminosavakat kódolnak; 2. kapcsolat van a kodonok második nukleotidja és a között, hogy a két nagy osztályba tartozó aminoacil-tRNS szintetázok közül melyik osztály tagja végzi a kodonhoz tartozó aminosav és a megfelelő tRNS összekapcsolását; 3. az egyes aminosavak moláris tömege és az őket kódoló kodonok száma negatív korrelációt mutat; 4. az egyes aminosavak fehérjékben való gyakorisága és az őket kódoló kodonok száma pozitív korrelációt mutat; 5. a sztenderd genetikai kód hibaminimalizálónak látszik (Haig és Hurst, 1991; Freeland és mtsai, 2003); és végül 6. a kodonok középső nukleotidja és a

kódolt aminosav katalitikus hatékonysága között összefüggés mutatkozik (Szathmáry és mtsai, 2007).

IV.2. A sztenderd genetikai kód hibaminimalizáló jellege

Az előző felsorolásban ötödikként szereplő tulajdonság azt jelenti, hogy a sztenderd genetikai kód, kodonkiosztása révén, a pontmutációs és misztranszlációs hibák következményeivel szemben ellenállónak látszik, ugyanis az adott kodonkiosztás a fenti események káros hatásait egy hipotetikus véletlen szinthez képest lecsökkenti (Freeland és mtsai, 2002). A pontmutációk elsődleges következménye, hogy hibás szekvenciájú mRNS kerül a riboszómára; misztranszlációkor pedig nem a riboszóma által aktuálisan beolvasott kodonnak megfelelő tRNS, és így nem a megfelelő aminosav érkezik a riboszómára. Végeredményben tehát mindkét hiba azt eredményezi, hogy a készülő fehérjének megváltozik az aminosavsorrendje, ez pedig bizonyos mértékben a fehérje funkcionalitásának rovására mehet. Fontos látnunk, hogy a beépülő helytelen aminosavhoz tartozó kodon a helyes kodontól egy pontmutáció esetén definíció szerint csak egy pozícióban különbözik, és a misztranszlációkor fellépő hibás kodon-antikodon kapcsolat létrejöttének is csak egy pozícióban való eltéréskor van értékelhető esélye, két pozíciónyi eltérés esetén ugyanis nem jön létre megfelelő erősségű kodon-antikodon kötés. Ez tehát azt jelenti, hogy a helyes és a helytelen aminosavak kodonja ezeknél a hibáknál majdnem teljes bizonyossággal csak egy pozícióban tér el.

A fehérjék aminosavsorrendjének megváltozásából eredő funkcióvesztés mértékét a fehérjeszerkezet és fehérjeműködés összetettségének következtében minden bizonnyal sok tényező határozza meg, mindezek közül azonban elsődleges szereppel bír, hogy a beépítendő helyes, és a beépített helytelen aminosav mennyire különbözik egymástól. Ennek a különbözőségnek a mérése és a megállapítása a tudományterület egyik kardinális kérdését jelenti, mely számos megközelítést szült (Koonin és Novozhilov, 2009). Az egyik legkorábbi ezek közül a Carl Woese által kidolgozott polaritási skála (*Polar Requirement Scale*), amely az egyes aminosavak között kísérletesen mérhető hidrofobicitásuk különbsége alapján állapít meg távolságokat (Haig és Hurst, 1991). Mivel az egyes aminosav-oldalláncok hidrofobicitása a

fehérjék térszerkezetét lényegesen meghatározza, a helyes térszerkezet pedig az optimális működés feltétele, adódik a következtetés, hogy a hidrofobicitás megváltozásának mértéke a funkcióvesztés súlyosságával jól korrelál (Freeland és mtsai, 2003). A Woese-féle polaritási skálához hasonlóan szintén az aminosavak fiziko-kémiai hasonlóságának becslésén alapul a Gilis-féle pontozó mátrix. Ez az egyes aminosavcseréket jelképező aminosav-párokhoz aszerint rendel hibaértékeket, hogy az adott csere mekkora változást idéz elő egy fehérje szabadentalpiájában (Koonin és mtsai, 2007). A bemutatott két módszertől alapvetően eltérő, ám az aminosavcserék hatásának becslésére ugyancsak alkalmazott módszert jelentenek az ún. PAM-mátrixok (*point accepted mutation*) (Freeland és mtsai, 2003), melyek empirikusan megállapított értékei az egyes aminosavcserék recens fehérjékben mérhető gyakoriságát, és így valószínűségét tükrözik, különböző számú (általában 75) generációra vonatkoztatva. Az egyes cserék gyakorisága pedig a PAM-mátrixok ilyen irányú alkalmazása mellett érvelők szerint negatívan korrelál az általuk okozott hiba súlyosságával, így ezek a mátrixok a cserék jelentőségének becslésére alkalmasak.

Mint fentebb rámutattam, egy pontmutáció vagy misztranszláció következményeként fellépő aminosav-csere esetén az eredeti és a csereként érkező aminosavak kodonja csak egy nukleotidpozícióban különbözik, vagyis ezen kodonok a mutációs térben egymás mellett helyezkednek el. Amennyiben egy hipotetikus kodonkiosztás olyan, hogy benne az egyes kodonpárok mutációs közelsége az általuk kódolt aminosav-párok valamely skála szerint mért közelségével arányos, úgy ez a kodonkiosztás a pontmutációból vagy misztranszlációból eredő aminosav-cserék hatását minimalizálja. A sztenderd kód hibaminimalizáló tulajdonsága tehát azt jelenti, hogy kodonkiosztása ilyen jellegű. A következőekben az e feltevés jogosságát, valamint a kapcsolódó kérdéseket vizsgáló főbb kutatásokat mutatom be.

IV.2.1. A hibaminimalizáló jelleg foka a sztenderd genetikai kódban

A fentebb bemutatott módszerek és elvek alapján láthatjuk, hogy egy adott kodonkiosztás hibaminimalizáló jellegének foka kvantifikálható. Ahhoz azonban,

hogy az így megkapott értéket értelmezni tudjuk, szükség van valamiféle háttérre, amelyhez képest a hibaminimalizáló jelleg viszonyítható.

Ezen megfontolások mentén végzete el David Haig és Laurence D. Hurst 1991-ben a kérdés első komoly számítási kapacitást hasznosító vizsgálatát (Haig és Hurst, 1991). Ennek során először egy, az egyes szinonim blokk véletlenszerű cserélgetését végző algoritmussal 10 000 darab véletlenszerű kódot generáltak, majd megvizsgálták, hogy az ezek által képzett halmazban a hibaminimalizáló jelleg mértéke milyen eloszlást mutat, és hogy a sztenderd genetikai kód ezen eloszláshoz képest milyen mértékben számít hatékony hibaminimalizálónak. Az elemzés során az egyes aminosav-cserék hatását elsősorban a fiziko-kémiai hasonlóság alapján becsülték meg, és azt kapták, hogy a sztenderd genetikai kódban a kodonok mutációs távolságával a legnagyobb korrelációt az aminosavak hidrofobicitás szerinti távolsága mutatja, ezen a téren pedig a sztenderd kód a véletlenszerűen generált kódok 99,98%-ánál jobban teljesít (Haig és Hurst, 1991). Egy további fontos eredményük az volt, hogy az első és a harmadik nukleotidpozíciót érintő mutációk esetén a másodikat érintőknél lényegesen nagyobb mértékűnek mutatkozik a hibaminimalizáltság foka, ebből pedig arra következtettek, hogy hibaminimalizáló jelleg elsősorban a transzlációkor fellépő hibák kivédését szolgálja. A replikációkor fellépő pontmutációk ugyanis a leolvasási keret, vagyis az egyes kodonokban a nukleotidok pozíciói által nem befolyásoltak, így az ilyenkor keletkező hibák kivédését szolgáló optimalizáció sem tehet különbséget a nukleotidpozíciók között (Haig és Hurst, 1991). Az eljárás későbbi, többek között éppen erre a megfigyelésre alapozó finomításai aztán még meggyőzőbb eredményeket hoztak: azt találták, hogy a sztenderd genetikai kód egy millió véletlenszerűen generált kód mindegyikénél jobban teljesít (Freeland és mtsai, 2003).

A vonatkozó kutatások valamivel kifinomultabb csoportját képezik azok a munkák, melyek a szinonim blokkok párokban való áthelyezésével véletlenszerűen generált kódok evolúcióját, vagyis a kodonkiosztás fokozatos megváltozását is megengedik. Koonin és Novozhilov egy ilyen elvekre épülő szimulációs vizsgálatban, a korábbi eredményekkel szemben arra jutottak, hogy egy átlagos random kód tipikusan 15-30

evolúciós lépés alatt a sztenderd genetikai kód által mutatott hibaminimalizáció szintjét eléri, az esetek jelentős részében pedig túl is szárnyalja azt (Koonin és mtsai, 2007). Egy további eredményük, hogy a kodonkiosztás változtatásával maga a sztenderd kód is tovább optimalizálható. Koonin és Novozhilov mindezek alapján levonja a következtetés, hogy a sztenderd kód közepes hibaminimalizációt jelentő kodonkiosztással bír, és így semmilyen szempontból nem tekinthető olyan mértékben optimalizáltnak, mint azt a korábbi eredmények sugallták (Koonin és mtsai, 2007).

IV.2.2. A kód evolúciója és a hibaminimalizáló jellegének eredetének kérdése

Mint láttuk, a sztenderd genetikai kód kodonkiosztása egyértelműen és határozottan nem véletlenszerű. Mik voltak tehát azok az okok és evolúciós események, amelyek ennek a specifikus mintázatnak a kialakulását hozták? A kérdés megválaszolását három fő megközelítés kísérli meg, melyek azonban nem kizárólagosak, és minden bizonnyal egyaránt hozzátesznek valamit a probléma tisztázásához (Koonin és Novozhilov, 2009).

A sztereokémiai hipotézis, melyet a kód eredetét tárgyaló fejezetben már röviden bemutattunk, központi állítása, hogy az egyes kodon/antikodon-aminosav asszociációk, és így a kódtábla szerkezete mögött az adott tripletek és aminosavak közötti direkt sztereokémiai affinitás áll (Yarus, 2009; Koonin és Novozhilov, 2009). Az elmélet helyessége azonban jelen állás szerint nehezen ítélhető meg, lévén csupán nyolc aminosavat illetőleg rendelkezünk releváns adatokkal (Yarus, 2009), a statisztikai analízis érvényessége pedig kétségbevonható (Koonin és Novozhilov, 2009). Egy további probléma a sztereokémiai hipotézis általános magyarázatként való elfogadásával, hogy nehezen látszik, hogy ez a mechanizmus hogyan eredményezhetné a sztenderd genetikai kód mintázatait. Mindazonáltal Yarus eredményei alapján valószínű, hogy a korai kodon-aminosav asszociációk kialakulásában szerepet játszott egy sztereokémiai alapú hatás, melyet aztán későbbi folyamatok többé-kevésbé kiegészítettek (Freeland és mtsai, 2003; Koonin és Novozhilov, 2009).

Szintén volt már szó a koevolúciós hipotézisről, melynek központi állítása, hogy a kód kiterjeszkedése során az új aminosavak kódba való integrációja a már kódolt aminosavak szintézisútvonalainak továbbalakulásával, módosulásával történt, mégpedig a fentebb már részletesen tárgyalt pretran-szintézis révén (Wong, 2005). Ez a mechanizmus pedig olyan kodonkiosztást eredményez, amelyben a rokon, és így fiziko-kémiai szempontból jó eséllyel hasonló aminosavak kodonjai is közel lesznek egymáshoz a mutációs térben. Mindez azt jelenti, hogy a koevolúciós hipotézis által a kód evolúciója során központinak tekintett folyamat a kodonkiosztás hibaminimalizáló jellegét a nélkül is képes előállítani, hogy konkrétan erre a jellegre történne szelekció (Wong, 2005; Koonin és Novozhilov, 2009).

Egy ilyen egyértelműen és határozottan jelenlévő tulajdonság esetén mindazonáltal érdemes megvizsgálni, hogy állhatnak-e adaptív folyamatok megjelenése mögött. Az adaptív hipotézis szerint a sztenderd genetikai kód kodonkiosztásának mintázata mögött aktív, a pontmutációkból és misztranszlációkból eredő káros hatások csökkentésére irányuló szelekció áll, amely a kodonkiosztás finomhangolása révén fejtette ki hatását (Koonin és mtsai, 2007; Koonin és mtsai 2009b).

A különböző megközelítések által lényegesnek tartott hatások viszonya és kizárólagossága érdekes kérdést jelent, a jelenlegi ismeretek alapján azonban csak erősen spekulatív megállapításokkal élhetünk ezzel kapcsolatban, a tudományterület prominensei között pedig egyelőre hiányzik az egyértelmű konszenzus. Koonin és Novozhilov egy kompromisszumos szcenárió mellett foglalnak állást (Koonin és Novozhilov, 2009), mely szerint valamennyire a sztereokémiai, és különösképpen a koevolúciós hipotézis felvetéseit is érdemes figyelembe venni, ám a sztenderd kód kodonkiosztását és így hibaminimalizáló jellegét ezek a hatások önmagukban nem magyarázzák. Véleményük szerint a sztereokémiai hatások legfeljebb a korai kodon-aminosav kapcsolat megjelenésére lehettek hatással, míg a kód szintézisutakkal összefüggő evolúcióját lényeges, ám nem kizárólagos tényezőként értékelik a hibaminimalizáló jelleg eredetének kérdésében. Mind Yarus, mind Wong szélsőségesebb álláspontón vannak. Yarus szerint a kodonkiosztás bizonyos hányada mögött egyértelműen a sztereokémiai hatások állnak (Yarus, 2009), Wong pedig

egyenesen odáig megy, hogy a koevolúciós hipotézis mellett a többi elmélet által javasolt hatásokat ugyan valósnak, de sok nagyságrenddel jelentéktelenebbnek értékeli (Wong, 2005).

Végül szükséges kitérnünk a sztenderd kód evolúcióját meghatározó tényezők és hatások kronologikus viszonyaira. A tripletek és aminosavak közti sztereokémiai affinitás szerepét, mint említettük, Koonin és Novozhilov (Koonin és Novozhilov, 2009), de csakúgy Freeland és munkatársai is (Freeland és mtsai, 2003) egy korai, az ősi genetikai kód megjelenésének idején érvényesülő hatásnak tartják, ebben a kérdésben tehát egyetértés mutatkozik. A kódtábla és a szintézisutak koevolúciójának, valamint a hibaminimalizáló jellegre való adaptációnak a viszonya kevésbé tekinthető tisztázottnak, leginkább a vonatkozó kutatások kis száma miatt. Koonin és munkatársainak 2009-es munkája (Koonin és mtsai, 2009) azonban érdekes elméleti felvetésekkel él. Ebben a korai aminosavak valószínűsíthető halmaza (lásd a kód kiterjeszkedését tárgyaló részt), valamint a sztenderd kód kodonkiosztása alapján, a lehető legkevesebb evolúciós lépést szükségességét szem előtt tartva megkísérelték az ősi genetikai kód kodonkiosztásának rekonstruálását, azt feltételezve, hogy ebben a stádiumban az egyes tripleteknek még csak az első két nukleotidja volt kódoló szerepű (8. ábra). Ezt követően pedig megvizsgálták, hogy az

	U	C	A	G
U	?/Leu	Ser	?	?
C	Leu	Pro	?	?
A	Ile	Thr	?	?/Ser
G	Val	Ala	Asp	Gly

8. ábra: A korai aminosavak halmaza és a jelenlegi kodonkiosztás alapján, a legkevesebb szükséges evolúciós változást feltételező módon rekonstruált hipotetikus ősi kód, melyben feltételezések alapján csak a kodonok első két nukleotidja volt kódoló szerepű. Az így kapott kód kiemelkedően, a sztenderd kódot felülmúlóan hibaminimalizáló jellegű. A kép forrása: Koonin és mtsai, 2009.

így kapott kodonkiosztás a hibaminimalizáció terén hogyan teljesít, és azt kapták, hogy a rekonstruált ősi genetikai kód ebben a tekintetben a jelenlegi kódot túlszárnyalja (Koonin és mtsai, 2009). Tekintve, hogy a csupán a korai aminosavakat

kódoló, tehát a kiterjeszkedés előtt álló hipotetikus ősi genetikai kódot vizsgálták, levonható a következtetés, hogy a genetikai kód és a szintézisutak ezt követő koevolúciója történetesen a hibaminimalizáló jelleg mértéknek csökkenését hozta.

V. Összefoglalás

A genetikai kód egy hozzárendelés, amely a genetikai információt hordozó nukleinsavak hármasaihoz a katalitikus feladatokat ellátó fehérjéket alkotó aminosavakat társítja. Ez a hozzárendelés a transzlációs apparátus tevékenysége során érvényesül, amikor is a DNS-ből származó információ alapján az egyes fehérjék szintézise végbemegy. Mivel azonban a DNS-től a fehérjék felé irányuló információáramlásban maguk a fehérjék is kulcsszerepet játszanak, a rendszer létrejöttének kérdése körkörös problémát jelent. Erre kínál megoldást, ha a rendszer eredetét az RNS-világban keressük. A transzlációs apparátus kialakulásán belül a genetikai kód megjelenésének kérdése egy kezdeti, részproblémát jelent. A kód megjelenésének adaptív vonása a kódoló koenzim fogantyú hipotézis szerint az aminosavak ribozimek általi katalitikus, koenzim-szerepű felhasználásban rejlett, mely során elsőként jelent meg az egyes aminosavak és nukleotid-szekvenciák közti asszociáció, vagyis maga a genetikai kód. A sztereokémiai hipotézis az egyes triplet-aminosav párok kialakulásánál a tripletet tartalmazó szekvencia és a triplet által kódolt aminosav közötti sztereokémiai affinitást jelöli meg vezérlő hatásként, amely az ősi genetikai kód kezdeti struktúráját kijelölte. A genetikai kód kialakulása során a következő lényegi fázist az egyre integráltabb és központibb szerepet kapó genetikai kód kiterjedése, vagyis a kezdeti néhány kódolt aminosav után újabbaknak a kódba való foglalása jelenti. A folyamat a koevolúciós hipotézis szerint a genetikai kód és az aminosavakat előállító biokémiai szintézisutak párhuzamos evolúciója révén ment végbe, amely során az új aminosavak megjelenése az egyes szintézisutak kiegészüléséhez fűződött. Az ilyen módon létrejött, sztenderd genetikai kód kodonkiosztása, mely az élővilágban közel univerzális, számos figyelemreméltó mintázattal bír. A prekurzor-termék aminosav párok elemzésével azonosíthatóak például a kódtáblában a bioszintetikus utak kiterjedésének nyomai. Egy további mintázatot jelent a blokkos szerkezet, amely az azonos aminosavat kódoló, szinonim kodonok csoportokba való rendeződését jelenti a mutációs térben. A kód ezen felül kimondottan hibaminimalizáló természetűnek mutatkozik, az esetleges pontmutációk és misztranszlációk miatt bekövetkező aminosavcserék hatását ugyanis a sztenderd kód kodonkiosztása csökkenti. Ennek a tulajdonságnak a hátterében egyrészt a koevolúciós hipotézis mechanizmusa, másrészt adaptív folyamatok állhatnak.

VI. Summary

The genetic code is a mapping, which associates the triplets of information carrying nucleic acids to amino acids, building blocks of proteins, whose main task is to carry out the catalytic function. This mapping is implemented by the activity of the translation apparatus, during which the synthesis of proteins based on the information derived from the DNA is being done. However, since in the flow of information from DNA to proteins, the latter play a crucial role, emergence and appearance of this system poses a seemingly tautological problem. This problem can be avoided if the origin of the system is hypothesized to have taken place in the RNA world. Within the emergence of the translation apparatus, the origin of the genetic code is an initial sub-problem. According to the coding coenzyme handle hypothesis, the adaptivity of the appearance of the genetic code was provided by the utilization of amino acids by ribozymes as coenzymes for catalytic purposes. This utilization means the first appearance of the association between amino acids, i.e. the appearance of the genetic code itself. The stereochemical hypothesis identifies the stereochemical affinity between nucleotide sequences contanining triplets and the amino acids coded by these triplets as the main governing force during the formation of amino acid-triplet pairs, setting out the initial structure of the primordial genetic code. The subsequent phase during the emergence of the genetic code was the expansion of the increasingly integrated and central code, during which the initially low number of coded amino acids grew. According to the coevolution theory, this growth occured via the paralell evolution of the genetic code and the biochemical synthesis pathways generating amino acids; the process of amino acid acquirement is thought to have happened by the extension of these pathways. The codon assignment of the thus formed genetic code, which in the sphere of life is nearly universal, displays a set of remarkable patterns. Upon analysing precursor-product pairs of amino acids, the footprints of the expansion of biosynthetic pathways can be detected. The block structure, which means that codons coding for the same amino acid are grouped in synonymous series in the mutational space, is another pattern. Moreover, the code seems expressly error minimizing: the function loss due to amino acid chages caused by point mutations and mistranslations is decreased by the codon

assignment of the genetic code. The origins of this might lie in the mechanism proposed by the coevolution theory, or in adaptive processes.

VII. Hivatkozások

Brooks, D., Fresco, J., Lesk, A., & Singh, M. (2002). Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. *Molecular Biology and Evolution*, 19 (10): 1645-1655.

Chen, X., Li, N., & Ellington, A. D. (2007). Ribozyme catalysis of metabolism in the RNA world. *Chemistry & Biodiversity*, 4(4): 633–55.

Crick, F. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology 12*, (1958): 138-163.

Freeland, S. J., Wu, T. A. O., & Keulmann, N. (2003). The case for an error minimizing standard genetic code. *Origins of Life and Evolution of Biospheres*, (4-5): 457-77.

Haig, D., & Hurst, L. (1991). A quantitative measure of error minimization in the genetic code. *Journal of Molecular Evolution*, 49(5): 708.

Joyce, G. F. (2002). The antiquity of RNA-based evolution. *Nature*, 418(6894): 214–21.

Koonin, E., & Novozhilov, A. (2009). Origin and evolution of the genetic code: the universal enigma. *IUBMB life*, 61(2): 99-111.

Kun Á, Pongor S, Jordán F, Szathmáry E (2007) Catalytic propensity of amino acids and the origins of the genetic code and proteins. In: Barbieri M. (ed), *The Codes of Life: The Rules of Macroevolution.* Springer, 39-58.

Kun Á. (2010). Az RNS világ és a hibaküszöb. *Magyar Tudomány*, 171(4): 388-395.

Kun Á. (2011). Az RNS Világ. *Természet Világa*, 142(10): 455

Miller, S. (1953). A production of amino acids under possible primitive earth conditions. *Science*, 117(3046): 528-529

Müller, M., Allison, J., & Hongdilokkul, N. (2013). Directed evolution of a model primordial enzyme provides insights into the development of the genetic code. *PLoS genetics*, 9(1): e1003187.

NCBI (2013). The Genetic Codes. Elérés: http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi. Elérés dátuma: 2013.05.15.

Novozhilov, A. S., & Koonin, E. V. (2009). Exceptional error minimization in putative primordial genetic codes. *Biology Direct*, 4: 44.

Novozhilov, A. S., Wolf, Y. I., & Koonin, E. V. (2007). Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. *Biology Direct*, 2: 24.

Szathmáry, E. (1993). Coding coenzyme handles: a hypothesis for the origin of the genetic code. *Proceedings of the National Academy of Sciences of the United States of America*, 90(21): 9916–20.

Szathmáry, E. (1999). The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends in Genetics*, 15(6): 223-9.

Trifonov, E. (2004). The triplet code from first principles. *Journal of Biomolecular Structure and Dynamics*, 22(1): 1-11.

Wochner, A., Attwater, J., Coulson, A., & Holliger, P. (2011). Ribozyme-catalyzed transcription of an active ribozyme. *Science*, 332(6026): 209-212.

Wong, J. T.-F. (2005). Coevolution theory of the genetic code at age thirty. *BioEssays*, 27(4): 416–25.

Yarus, M., Widmann, J. J., & Knight, R. (2009). RNA-amino acid binding: a stereochemical era for the genetic code. *Journal of Molecular Evolution*, 69(5): 406–29.

VIII. Köszönetnyilvánítás

Legfőképpen köszönettel tartozom témavezetőmnek, Dr. Kun Ádámnak a dolgozat elkészítésében nyújtott segítségért és a téma mélyebb megértését és átgondolását lehetővé tevő beszélgetésekért.

Szeretnék továbbá köszönetet mondani Édesanyámnak, Dr. Boros Juditnak, segítségéért és kitartó támogatásáért.

NYILATKOZAT

Név: MÁRKUS BENJÁMIN

Neptun azonosító: U1U87M

ELTE Természettudományi Kar, **biológia alapszak**

Szakdolgozat címe: Az élet keletkezése - A genetikai kód eredete

A szakdolgozat szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló munkám eredménye, saját szellemi termékem, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam.

Tudomásul veszem, hogy plágiumnak számít:

- szó szerinti idézet közlése idézőjel és hivatkozás megjelölése nélkül;
- tartalmi hivatkozás a forrás megjelölése nélkül;
- más személy publikált gondolatainak saját gondolatként való feltüntetése.

Kijelentem továbbá, hogy a szakdolgozat leadott nyomtatott példányai és elektronikus változata szövegükben, tartalmukban megegyeznek.

Budapest, 2013.05.23.

a hallgató aláírása