

OTDK

Egy kiválasztott, magányos töltött alfa-hélixet tartalmazó fehérjecsalád evolúciójának
elemzése

Dobson László

Témavezető: Dr. Gáspári Zoltán

Pázmány Péter Katolikus Egyetem

Információs Technológiai Kar

Konzulens: Dr. Gáspári Zoltán



Pázmány Péter Katolikus Egyetem

Információs Technológiai Kar

Budapest

2012.

Tartalomjegyzék

1. Rezümé	4
2. Abstract.....	5
3. Bevezető.....	6
4. Célkitűzés.....	7
5. Előzmények, irodalmi áttekintés.....	8
5.1. Motívumok	8
5.1.1 Coiled coil.....	8
5.1.2 Magányos töltött alfa-helix	10
5.1.3 Funkcionálisan rendezetlen szakaszok	10
5.2. A NONO_HUMAN	11
6. Tervezés részletes leírása, felhasznált módszerek	12
6.1. Adatbázisok.....	12
6.2. Bioinformatikai algoritmusok	13
6.2.1. BLAST	13
6.2.2 Többszörös illesztés	13
6.2.3 Filogenetikus fa.....	14
6.3. Egyéb bioinformatikai eszközök.....	14
6.3.1 RasMol.....	14
6.3.2 Coiled coil szakaszok keresése és vizsgálata.....	14
6.3.3 Magányos töltött alfa-helix szakaszok keresése	15
6.3.4 Rendezetlen szakaszok keresése	15
6.3.5 Domének keresése	15
6.3.6 Exon-intron határok keresése.....	15

7. Elemzés, értékelés.....	16
8. Összefoglalás, kitekintés.....	24
9. Irodalomjegyzék	25
10. Köszönetnyilvánítás.....	27

1. Rezümé

A fehérjeevolúció során a térszerkezet egyes elemei vagy akár egésze is megváltozhat. Korábbi eredmények alapján valószínűsíthető, hogy három speciális térszerkezeti elem, a charged single alfa-helix, a coiled coil és a rendezetlen szakaszok viszonylag könnyen egymásba alakulhatnak. A célom konkrét példákon megvizsgálni az ilyen átalakulások lehetőségét, majd az érintett fehérjék listájának bővítésével átfogó képet alkotni a változások mikéntjéről és az azokat kísérő egyéb evolúciós folyamatokról.

A dolgozatban egy, a fent említett struktúrákkal várhatóan rendelkező fehérjét (NONO Human(Non-POU domain-containing octamer-binding protein)/splQ15233) vizsgáltam különböző bioinformatikai szempontok szerint. Többféle módszerrel kerestem térszerkezeti elemeket, illetve ezek mellett összehasonlítást végeztem más fehérjékkel, ahol azt figyeltem, mely konformációk jelennek meg több helyen és hogy ezek milyen összefüggésben vannak egymással és egyéb doménnel. Az ezekből kapott eredmények alapján filogenetikus fát készítettem, mely alapján evolúciós elemzéseket végeztem. Megkerestem a fehérjét kódoló mRNS-variánsokat és vizsgáltam mely szakaszok kódolják a fehérjét, és melyek vágódnak ki a splicing során.

A predikciókat összevetettem a szakirodalomban olvasottakkal és a már máshol fellelhető részadatokkal. A coiled coil szakaszoknál az aminosav sorrend nagyrészt egyezik, így elmondható, hogy a fehérjéken belül egy erősen konzervált részt sikerült találni. A vizsgált fehérjék nagy része a coiled coil/charged single alfa-helix szakaszon kívül mutat csak jelentős eltéréseket.

Megállapítható, hogy a vizsgált fehérjéken a coiled coil szakasz minden esetben együtt jelent meg két RRM (RNA recognition motif) régióval, amelyek így együtt a DBHS (Drosophila behaviour, human splicing) domént alkotják. További vizsgálatok illetve a friss szakirodalom rávilágítottak, hogy a coiled coil szerkezet a DBHS egyik alkotóeleme.

A látottak megerősítették, hogy térszerkezetek könnyen keresztbe-jósolhatóak, nincs egységes és egyértelmű algoritmus annak eldöntésére, hogy melyikkel állunk szemben. A vizsgálatokat több adatbázis fehérjéin is elvégeztem, ez a részletesebben annotált fehérjéknél megerősítette a már korábban dokumentált eredményeket, és a nem annotált, de az adatbázisokban fellelhető BLAST találatok esetén is jó egyezést kaptam. A jóslások alapján feltételezhető, hogy a vizsgálatok során talált, még nem annotált fehérjék a NONO fehérjecsalád ismert tagjait egészítették.

2. Abstract

Protein evolution often leads into the alteration of some part or the whole spatial structure. Previous results indicate, that three special conformation, charged single alfa-helix, coiled coil and instrinsically disordered proteins can be relatively easily converted into each other. The purpose of this thesis to study the possibility of variations on specific examples, expand the list of involved proteins and making a broad evaluation of the spatial changes and the evolutionary processes.

In this thesis, a protein expected to have one or more of the structures mentioned above was examined from a number of different aspects. Several methods were used to examine spatial orders in NONO_HUMAN (Non-POU domain-containing octamer-binding protein) and proteins besides it to figure out which conformations appear in several places and how they are related to each other or to other domains. mRNA variants able to encode the proteins were also identified either the relationship of the exon structure, the protein domains and structural motifs were examined.

Comparison was made between results, articles and other datas and phylogenetic tree was made based on them for evolutionary analysis.

At the coiled coil segments the amino acid sequences are almost the same, thus we can declare that strongly conserved sections were found. Most of the examined proteins show variance only outside the coiled coil/single charged alfa-helix parts.

In all of the examined proteins the coiled coil region appeared with two RRM (RNA Recognition Motif) region and together they form the DBHS (Drosophila behaviour, human splicing) domain. Our results are in the line with recent articles pointing out, that the coiled coil (along with the RRM domains) is a component of the DBHS.

Results confirmed that coiled coils and charged single alfa-helices can be easily cross-predicted, there aren't clear algorithms to decide what alterations were seen. Similarity can be noticed at disordered segments too, however it's not nearly as clear as in the case of other two conformations. The tests were also made on other homologous and confirmed already documented results on richly annotated proteins. In some places already gathered informations were expanded with newer ones. The results assume, that proteins were found without annotation expand the NONO family with variants of other species.

3. Bevezetés

A fehérjék alapvető fontosságú makromolekulák, fontos szerepüket jelzi, hogy minden sejtben lejátszódó folyamatban részt vesznek. Számos feladatot láthatnak el: lehetnek a sejt szerkezeti komponensei, elláthatnak védőfunkciót, lehetnek transzportproteinek, jelátviteli fehérjék, szabályozó fehérjék, illetve számos esetben enzimek, azaz valamilyen biokémiai folyamat katalizátoraként segítik elő a sejt életben maradását. A fehérjék funkcióit nagymértékben befolyásolja térszerkezetük, melynek vizsgálatára többféle módszer született az elmúlt évtizedekben

Az élettudományokban az elmúlt 30 évben egyre nagyobb teret kapott a számítástechnika és az informatika. Kezdetben a bioinformatika nem alkotott önálló tudományágat, csak az adatok összefoglalásában, elemzésében segítették az experimentális módszereket alkalmazó szakembereket az új lehetőségek.

A bioinformatika a biológiai információ kezelésére képes számítógépes algoritmusok alkalmazásával foglalkozik, melyek az alkalmazott matematika, a számítástechnika és a statisztika kombinálásával jöhettek létre. Ezek a módszerek a biológiai információ nagyléptékű feldolgozására jöttek létre. A bioinformatika feladata az adatbázisok kezelése, az ott felgyűlő információ kezelése és biológiai értelmezése. A bioinformatika jelenleg nélkülözhetetlen kiegészítője számos kísérleti módszernek (pl.: újgenerációs szekvenálás), illetve sok esetben képes megbízható jóslatokat tenni, csökkentve ezzel az elvégezendő kísérletek számát.

Ma a bioinformatika képes kvalitatív válaszokra olyan területeken is, mint például a szekvencia-analízis, struktúra meghatározás röntgen-krisztallográfia és NMR adatok alapján, strukturális domének azonosítása, biomolekulák interakcióinak vizsgálata és szimulációja, vagy akár konkrét (biológiailag hatékony) molekulák tervezése.

A bioinformatika jól használható olyan esetekben is, mint a klaszterezés, szekvencia-illesztés, struktúra/funkció összefüggések megállapítása, molekuláris vizualizáció, teoretikus validáció, biológiai és evolúciós modellek prediktív erejének becslése.

A dolgozat célja, hogy a humán NONO-n (Non-POU domain-containing octamer-binding), mint modellfehérjén és annak homológjain a coiled coil, a magányos töltött alfa-helix (charged single alfa-helix (a továbbiakban CSAH)) és a rendezetlen szerkezetű struktúrákat vizsgálja bioinformatikai módszerekkel, az eredményeket vesse össze az adatbázisok eredményeivel, illetve ezek alapján elemezze az egymásba alakulás lehetőségét és a lehetséges evolúciós útvonalakat.

A coiled coil és a charged single alfa-helix szerkezeti motívumok számos fehérjében előfordulnak és jelentős szerepet játszanak azok strukturális kialakításában, a kettő közötti határvonal azonban nem éles.

4. Célkitűzés

Feladatomban egy tetszőleges magányos töltött alfa-hélixet (CSAH) tartalmazó fehérje vizsgálata volt, a biológiai jelentőség áttekintése, különös tekintettel a más fehérjékkel kölcsönhatásban álló szakaszaira. A fehérje homológjainak azonosítása az UniProt adatbázisban, humán és néhány kiválasztott további genomban. Kimerítő predikciók végzése a fehérjékben található rendezetlen, coiled-coil és CSAH szakaszok azonosítása céljából. A különbözőnek jósolt szakaszok közötti átfedéseket elemzése. Néhány kiválasztott fehérje kódoló szakaszának ismeretében az esetlegesen azonosított térszerkezeti átmenetek mutációs hátterének értelmezése.

A vizsgálatokat a Non-POU domain-containing octamer-binding protein (NONO_HUMAN/Q15233) fehérjén végeztem. Szerkezetileg azért jelentős a vizsgálat szempontjából, mert a témával kapcsolatos cikkek alapján várhatóan tartalmaz coiled coil és CSAH régiót^{1,2}, illetve számtalan fontos kölcsönhatásban szerepet játszhat, amelyet biokémiai és terápiásan is alkalmazni lehet³, viszont eddig nem végeztek rajta átfogó bioinformatikai vizsgálatokat.

Először homológ szekvenciapárokat kerestem, majd a találatokon többszörös illesztést hajtottam végre. A kereséskor több adatbázist felhasználtam, és csak a nagymértékben egyező aminosav-láncokat vizsgáltam tovább. Mindegyik fehérjén végeztem coiled coil, CSAH és rendezetlen szakasz keresést, illetve megvizsgáltam milyen egyéb doménekkel rendelkezhetnek. Figyelembe vettem a szekvencia konfliktusokat, hogy ezek befolyásolják-e a kapott eredményeket.

Kerestem a NONO_HUMAN-t lekódolni képes mRNS-variánsokat és ezek exon-intron határait.

A fenti eredmények felhasználásával filogenetikus fát rajzoltattam, amely segítségével levontam az evolúciós következtetéseket.

A feladatokhoz különböző programokat vagy saját magam illetve mások által írt scripteket és programokat használtam fel. Ezek python és c++ nyelven íródtak. Egyes feladatok elvégzéséhez az R statisztikai programcsomagot használtam, ennek alapszintű megismerése szintén része volt feladataimnak. A grafikai vizualizációra a RasMol nevű programot használtam.

5. Előzmények, irodalmi áttekintés

A fehérjék esetén több, egymásra épülő strukturális szintet különböztethetünk meg. Térszerkezetük kialakulása elsősorban az őket felépítő aminosav szekvencia által meghatározott (elsődleges szerkezet). Másodlagos struktúra alatt a polipeptidlánc lokális konformációját értjük. Az α -helikális vagy β -redőzött és egyéb másodlagos szerkezeti elemek együtt hozzák létre a harmadlagos szerkezetet. Negyedleges struktúra azokat a fehérjéket jellemzi, amelyeket több alegység (peptidlánc) alkot.

Jelen dolgozat a speciális térszerkezeti elemekre fókuszál, azon belül is a coiled coil, CSAH, illetve a rendezetlen részletekre. Ezeket a konformációkat elsősorban az aminosav sorrendből tudjuk jósolni. A coiled coil, a CSAH illetve a rendezetlen részek könnyen összekeverhetőek egymással⁴, jelenleg még nem létezik olyan egységes algoritmus ami a motívumokat egyértelműen megtalálná⁵, ezért sok a keresztbejósolás. Nagyfokú változékonyságuknak következtében könnyen egymásba alakulnak feltehetően mind a fehérjék belső dinamikája során, mind evolúciós értelemben, az egyik vagy másik szerkezetet előnyben részesítő aminosav-helyettesítések révén. Ennek ellenére mostani tudásunk már lehetővé teszi ezeknek a térszerkezeteknek az újrarendszerezését. Sense mutation esetén ugyan változik a DNS nukleotid sorrend, azonban ugyanazt a fehérjét fogja kódolni, így nem történik funkcióbeli változás. Missense esetén mutálódik a DNS, de kémiaiailag hasonló, indifferens aminosav épül be, ezért nincs fehérjeszinten jelentős biológiai változás. Nonsense mutáció esetén stop-kodonra mutálódik a DNS, ilyenkor teljes funkcióvesztés következhet a hiányzó rész miatt⁶. A lehetséges evolúciós utak tervezése során figyelembe kell venni, hogy az egyes változások mutációk által elérhetőek legyenek, miközben a funkcionalitás bizonyos szinten marad. Az alábbiakban egy adott fehérje struktúráját vizsgálom, és az aminosav módosulás hatására bekövetkező funkció és térszerkezetbeli változásokat figyelem.

5.1. Motívumok

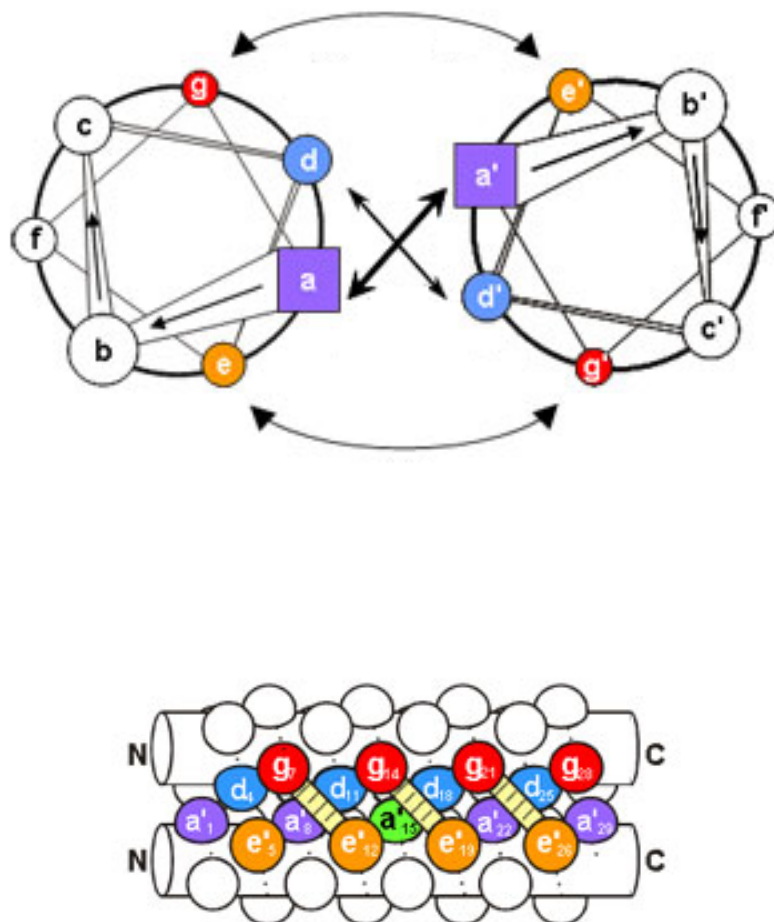
5.1.1. Coiled coil

A coiled coil egy strukturált szerkezeti motívum, amelynek fontos szerepe van például a több alegységű fehérjék kialakításában. Szerepét tekintve nagyon sokoldalú, ezért fehérje tervezés szempontjából ideális jelölt. Molekuláris „tépőzárnak” is hívják, mivel a sejten belüli struktúrák és polipeptid láncok összetartásában is szerepet játszik. Coiled coil motívumok számos biológiai kontextusban előfordulnak, pl. izomfehérjék, transzkripciós faktorok, sejtváza fehérjék, sejtfelszíni- és vírus-felszíni fehérjék, tumor szupresszorok, molekuláris motorfehérjék, és sejtváza komponensek (részletesebben: 1. táblázat). A szekvenciaadatok alapján az összes eukarióta fehérje 10%-ában van jelen (<5% a prokariótákban).⁷

Felépítését tekintve (legalább) két egymásra tekeredett alfa-hélix alkotja. Az oldalláncok állása többféle lehet, merőleges, párhuzamos, vagy a kettő közötti. A stabilitás fenntartását inter/intrahelikális

kölcsönhatások segítik, a két jobbra tekeredő párhuzamos alfa-helix egy balra tekeredő szupertekercsben helyezkedik el (1. ábra).

Ezt az elrendezést gyakran hívják „knobs into holes”-nak. Ilyenkor a helix oldallánca 4 másik oldalláncba van csomagolva. Tipikus példa klasszikus coiled coil-ra a leucine zipper, amely több transzkripció regulátor fehérjében megtalálható. A coiled coil-okat általában rendezetlennek tekintik monomer formában és rendezettnek szupercoil formában. A kettő közötti átalakulást egy monomer peptid lánc és egy csomagolt coiled coil dimer/oligomer közötti kétállapotú átmenet írja le (újabb eredmények szerint sok esetben ehhez szükség van egy úgynevezett katalizátor szekvenciára is). A klasszikus 2 szálú formán túl bonyolultabb coiled coil alakzatok is létrejöhetnek. A főbb szempontok a tengelyes szimmetria, a hidrofób mag valamint a töltött ionpár jelenléte⁷.



1. ábra: Kétszálú coiled coil felül- és oldanézetből: „a” és „d” hidrofób magok, „e” és „g” töltött ion pár. Jól látható a 7 aminosavból álló ismétlődés⁸

Funkció	példák
Transzkripció faktor	NONO [Q15233]
Motorfehérje	Kinesin-1 heavy chain [P33176]
Izomkontrakció	Myosin [P11055]
Adhéziós fehérjék lokalizációjának regulációja	Angiomotin-like protein 2 [Q9Y2J4]
Struktúrfehérje (csillók)	Rootletin [Q8CJ40]
Sejtprolifерáció és differenciáció	Coiled-coil alpha-helical rod protein 1 [Q8TD31]
Jelátviteli út inicializáció	Dixin [Q80Y83]
Sejtmag transzport: receptor aktiváció	Calcium-binding and coiled-coil domén-containing protein 1 [Q8CGU1]
Splicing faktor	SFPQ protein [P23246]

1. táblázat: Néhány jellegzetes coiled coil funkció, illetve annak előfordulása (zárójelben az UniProt azonosító)⁹

5.1.2. Magányos töltött alfa-hélix

A magányos töltött alfa-hélix (CSAH) egy nemrégiben felfedezett töltéssel rendelkező egyláncú alfa-helikális struktúra kis komplexitással, sok ismétléssel. A benne található pozitív/negatív töltésű aminosavak alapján szerkezetileg könnyen jósolható, hogy a szekvencián belül hol lesz alfa-helix. Akkor beszélhetünk CSAH-ról, ha az i , $i+3$ vagy $i+4$ helyen váltakozó töltésű aminosavak helyezkednek el. Ez az egyébként ritka motívum viszonylag gyakori az emberi fehérjékben. A szerepe lehet mechanikai (pl. miozin VI előkar), illetve sokszor fordul elő DNS/RNS kötésben, génexpresszióban részvevő fehérjében. Ugyan tudjuk, hogy milyen funkciójú fehérjékben szerepelnek, azonban az itt betöltött pontos szerepük egyelőre ismeretlen².

5.1.3. Funkcionálisan rendezetlen szakaszok

A rendezetlen fehérjék/szakaszok nem rendelkeznek időátlagban állandó átlagos harmadlagos szerkezettel, hanem számos különböző konformáció közötti gyors, folyamatos átalakulásban vannak. Fontos, hogy ezek a szakaszok ilyen formában fejtik ki biológiai funkciójukat, szekvenciákból való felismerésükre számos különböző algoritmus született. A fehérjék hasonlóságát a rendezettséggel szokták párhuzamba állítani. Fehérjéket akkor nevezünk ugyanúgy rendezettnek, ha ugyanazzal a fő másodlagos szerkezettel, ugyanolyan

orientációval és kapcsolódással rendelkeznek. Ez azonban egy komoly szubjektív tényezőt hagy nyitva, miszerint mit tartunk „fő” másodlagos szerkezetnek. Sok esetben a kutatók sem tudnak megállapodni egyes fehérjék besorolásán, mivel szerkezeti tényezők mellett/helyett gyakran evolúciós és funkcionális tényezőket is figyelembe vesznek¹⁰.

5.2 A NONO_HUMAN

A humán NONO (Non-POU domén-containing octamer-binding protein) egy 471 aminosav hosszú fehérje, funkcióit a lenti táblázatban foglaltam össze (2. táblázat). A szerkezetére azért figyeltem fel, mert a témával kapcsolatos cikkek szerint coiled coil régiót tartalmaz, de mivel az IDP/CC/CSAH keresztpredikció gyakori, meg akartam vizsgálni, hogy valójában milyen régiókból is áll¹. A sejtmagban fordul elő, megjelenését a szív, a tüdő, az agy, a vese, a harántcsíkolt izom, a hasnyálmirigy, a méhlepény és a máj szövetében is kimutatták, tehát nem mutat jelentős szövet-specifikusságot génexpresszió szempontjából¹.

<i>Funkció</i>	<i>Mechanizmus</i>
DNS repair	SFPQ-NONO complex heterodimer és a MATR3 fehérje interakciójával
DNS rekombináció	SFPQ-NONO complex heterodimer formában a topoizomeráz I/TOP1 funkciójának modulációja
DNS repair	SFPQ-NONO complex heterodimer direkt kötődik a DNS-hez, és kooperál a Ku70/G22P1-Ku80/XRCC5 dimerrel
Transzkripció aktiváció	NONO direkt módon kötődik a célgén enhancer elemének IAP (intracisternal A particle) régiójához, aktiválva a transzkripciót
Transzkripció szabályozása	The SFPQ-NONO-NR5A1 complex kötődik a CYP17 promoteréhez, cAMP-függő transzkripcionális aktivitást mutat
mRNS érés	
mRNS splicing	SFPQ fehérjével heterodimert alkotva
Paraspeckle formálás	PSPC1-el

2. táblázat: A NONO_HUMAN szerepe^{11,12}

6. A tervezés részletes leírása, felhasznált módszerek

6.1. Adatbázisok

A bioinformatikában felhasznált információkat adatbázisokban tárolják, amelyek folyamatosan bővülnek a különböző kísérleteket végző kutatók jóvoltából. Adatbázisokra azért van szükség, mert a feldolgozatlan DNS és fehérje szekvencia (egyelőre) sokszorosan meghaladja az átlátható és más módon rendszerezhető mennyiséget. Az adatbázisokban található adatok részletes kiértékelése folyamatosan történik, azonban az ott található információ sokszor hiányos, vagy egyes esetekben téves is lehet. A különböző adatbázisok tartalma különböző lehet, pl.: fehérje-adatbázis (UniProt). A különböző adatbázisok kapcsolatban állhatnak egymással, sokszor egymás adatait is felhasználják.

Az NCBI (The National Center for Biotechnology Information) az első nyilvános adatbázis, amely az adatok tárolása mellett a számítógépes biológiai kutatásokban (szekvenciaanalizáló szoftverek fejlesztése) is tevékenyen részt vesz. Az NCBI alá tartozik többek között az Entrez univerzális keresője, a PubMed (többmilliós kereshető adatbázis cikkek összefoglalói között). Az NCBI a GenBank (DNS) és a RefSeq (mRNS, fehérje) szekvenciaadatbázisára épül.

Az UniProt a legrészletesebben annotált fehérje adatbázis, ahol többféle módon is rászűrhetünk fehérjékre vagy szekvenciákra. Az egyes lekéréseknél TrEMBL (feldolgozatlan adatokat tárol, esetleges automatizált annotációkkal) és a SwissProt (részletesen annotált fehérjék, melyekhez az információt szakemberek által lefuttatott vizsgálatok biztosítják) adatbázisokra támaszkodik. Az egyes fehérjékről részletes információt kapunk, úgy, mint eredet, aminosav sorrend (és az azon belül fellelhető struktúrák), biológiai funkció, ontológia, kapcsolódó szakirodalom. Egyes DNS vagy fehérjeláncok esetében a különbözőségi szekvenálási kísérletek és projektek hatására többféle szekvencia is adott lehet egyféle azonosítóval rendelkező bejegyzéshez. Ezek az UniProt-ban szekvenciakonfliktusként vannak feltüntetve.

Mindkét adatbázis beépített BLAST funkcióval is rendelkezik.

6.2. Bioinformatikai algoritmusok

6.2.1. BLAST (Basic Local Alignment Search Tool)

Ez az algoritmus adott hosszúságú, adott értéknél nagyobb hasonlóságú (de nem feltétlenül azonos) szegmenspárokat keres két szekvencia között. Találat esetén mindkét irányba növeli a szegmenspárt egy bizonyos küszöbértékig. A hasonlóságok pontozását az átmeneti valószínűségeket leíró mátrixokkal adjuk meg, amelyek figyelembe veszik, hogy az egyes aminosavak milyen valószínűséggel mikre cserélődnek ismert esetekben. Többféle pontozás létezik, az „engedékenyebbeknél”, magasabb lehet a zaj mértéke (azaz nem rokon fehérjék is találatként fognak szerepelni), a szigorúbbaknál viszont előfordulhat, hogy a megkötések miatt egyes rokon fehérjék is kimaradnak. A pontozást a statisztikai szignifikanciánál használjuk fel, mely két értéket veszünk figyelembe: a P-value annak a valószínűsége, hogy a kapott összerendezés a véletlen szüleménye. Az E-value a várható gyakoriság: adatbázison való kereséskor a véletlennek köszönhetően várható, a tényleges találattal azonos pontszámmal rendelkező találatok száma. Az E-értéket növeli a keresőbe vitt szekvencia hossza (hosszabb szekvencia esetén nagyobb az esély a lokális, véletlen azonosságokra), illetve hogy az adatbázis mérete (nagyobb adatbázisban könnyebb találni hasonlót, ami csak a véletlennek köszönhetően tűnik hasonlónak). Mind a P-, mind az E-value esetében a kisebb érték a jó, tipikusan $E=1e-10$ alatt tartjuk jónak a találatokat. Attól függően, hogy milyen szekvenciákat keresünk, különböző BLAST programok állnak rendelkezésre.

A dolgozatban meghatározott eredményekhez a blastp-t használtam, mely fehérje-fehérje hasonlósági keresést végez. A kiértékelt adatoknál BLOSUM 62-es mátrixot használtam, azokat a találatokat vettem figyelembe, ahol a kimenetben 0 volt a E-value numerikus értéke.

6.2.2. Többszörös illesztés

Szekvencia-illesztés esetén a hasonlóság mértéket keressük kettő vagy több szekvencia között. A nagy hasonlóság rokonságra utal. Az összerendezés során hézagokat is szűrhatunk be, így az egyezések száma nőni fog. Ebben az esetben akár 100%-os azonosságot is elérhetnénk, ami biológiailag értelmetlen, ezért a végeredményt pontozzuk az egyezések, a hézagok és az eltérések mértékében. A pontozáshoz itt is hasonlósági mátrixokat használunk, bár más paraméterekkel.

A ClustalX¹³ egy többszörös-illesztést végző program, amely az eredményt grafikusán is megjeleníti. Az algoritmus először minden bemenő szekvenciát összerendez az összes lehetséges módon, majd ezek alapján távolságokat számít a szekvenciapárok között, majd az ennek alapján megállapított sorrend szerint illeszti a szekvenciákat.

Tehát első esetben, a BLAST analízissel egy szekvenciát hasonlítottunk össze az adatbázisban található összesel, míg utána a Clustal segítségével a már megszerzett találatokat hasonlítottuk össze egymással.

6.2.3. Filogenetikus fa

A molekuláris filogenetika a fajok, organizmusok vagy géncsaládok közötti kapcsolatok felderítésével foglalkozik, ami segítséget ad az egyes (az adott fehérjével/DNS szekvenciával rendelkező) fajok közötti evolúciós rokonság meghatározásában. A filogenetikus fa az adott szekvenciák közötti kapcsolatokat ábrázolja grafikusán. Elkészítésekor már feltételezzük, hogy a vizsgált szekvenciák között evolúciós kapcsolat van. A nehézséget ugyanaz okozza, ami a többszörös illesztés esetében is: a szubsztitúciók, az inszerciók és a deléciók. Egy fát lehet gyökérrel vagy anélkül ábrázolni, előbbi az elágazásokat időrendben mutatja. Ez általában akkor lehetséges, ha rendelkezünk információval a szekvenciák időbeli megjelenéséről, vagy, ha a vizsgált adatok között van olyan, amelyik egyértelműen a vizsgált csoporttól korábban elkülönült szekvencia (úgynevezett outgroup). A legmegfelelőbb fa kiválasztásához az összes lehetségeset meg kell vizsgálni. A törzsfa már a többszörös illesztés alapján készül. A fa ábrázolásához oldalon található BioNJ alkalmazást választottam¹⁴.

6.3. Egyéb bioinformatikai eszközök

6.3.1. RasMol

A RasMol programmal fehérjéket, nukleinsavakat és kisebb molekulákat lehet grafikusán ábrázolni különböző szempontok szerint. A program a beolvasott koordinátákat tartalmazó fájlból a fehérje vagy nukleinsav 3-dimenziós modelljét többféle módon (szálak, felület stb.) képes megjeleníteni, illetve lehetőséget ad néhány alapvető számításra (pl.: atomok távolsága). Az adatokat tartalmazó fájlt a Protein Data Bank oldaláról¹⁵ töltöttem le.

6.3.2. Coiled coil szakaszok keresése és vizsgálata

Coiled coil szakaszok számítógéppel történő meghatározása korábban Fourier transzformáció és profil alapú módszerek alapján történt. Mivel ezek gyakran teljesen figyelmen kívül hagyták a jellegzetes 7 aminosavból álló ismétlődéseket, az újabb algoritmusok már (részben) ezek felismerésén alapulnak. Coiled coil szakasz keresésekor különböző méretű ablakolások lehetségesek¹⁶.

Coiled Coil kereséshez a Svájci Bioinformatikai Intézet Coils programját használtam¹⁷, amely 3 méretű ablakolással, szöveges és grafikai kimeneti formátummal is rendelkezik.

Az R egy statisztikai számításokra és grafikus megjelenítésre alkalmas programcsomag¹⁸. Számok kiterjesztéssel rendelkezik, melyek között több bioinformatikai is van. A vizsgálatok során a dolgozatban

szereplő eredményekhez a ProCoil addont használtam, mely adott coiled coil szakasz szálainak számát jósolja. Az eredményeket később megerősítette a LOGICOIL algoritmus is, mely egyedülálló módon pusztán a szekvencia alapján, külső adatbázis nélkül prediktál oligomerikus állapotot.¹⁹

A RasMol programmal történő vizualizációnál a coiled coil szakaszt a SOCKET szerver -a térszerkezetet figyelembe vevő- predikciója alapján jelöltem²⁰.

A felhalmozódó információt saját készítésű cpp programokkal értékeltem ki.

6.3.3. Magányos töltött alfa-hélix szakaszok keresése

CSAH szakaszok kereséséhez az témavezetőm közreműködésével korábban fejlesztett szerverprogramot használtam. Kétféle módszer létezik CSAH detekcióra: A SCAN4CSAH az CSAH-ot stabilizáló ionikus kölcsönhatásokat előnyben részesítő pontozáson alapszik, a FT_CHARGE az ismétlődő töltésmintázatot ismeri fel Fourier transzformáció segítségével. Az eredményekhez a mindkét módszer alapján megtalálható konszenzus szekvenciákat vettem figyelembe²¹.

6.3.4 Rendezetlen szakaszok keresése

Rendezetlen szakaszok keresését az Enzimológiai Intézet²² honlapján található programmal kerestem. Az algoritmus az aminosavszekvencia alapján kalkulált páronkénti energiatartalom alapján számol, feltételezve, hogy az aminosav-összetétel nem biztosít elegendő energiát kedvező kölcsönhatások kialakítására.

6.3.5 Domének keresése

A különböző régiókat a Pongor Sándor csoportja által fejlesztett és fenntartott, az ICGEB (International Centre for Genetic Engineering and Biotechnology) oldalán elérhető, SBASE domén kereső programmal kerestem²³. A különböző doméneket saját (python) programokkal is kerestem a teljes SwissProt adatbázison.

6.3.6. Exon-intron határok keresése

Ezeknek a határoknak vizsgálata azért fontos, hogy megtudjuk, a vizsgált térbeli struktúrák hány részből állnak. Az mRNS szakaszokat az NCBI adatbázisán kerestem, ahol jelölve voltak az exonhatárok. A transzlációt "szimuláló" algoritmusok 3-as ablakolással működnek (így a 3 nukleotidból álló kodon összes lehetséges kódolását figyelembe veszik), ami oda-vissza lefuttatva mRNS variánsenként 6 lehetséges polipeptidláncot ad eredményül, melyek közül 1 kódolja a keresett fehérjét. A megfelelő szakaszok fordításához az Európai Bioinformatikai Intézet által közzétett EMBOSS programcsomagot használtam²⁴

7. Elemzés, értékelés

Először hasonlósági keresést végeztem az UniProt adatbázisban a NONO_HUMAN fehérjén. A találatok közül csak azokat vettem figyelembe, melyeknél az E-value numerikus értéke 0 volt. Az illesztés után a fehérjék jelentős része nagymértékben egyezett, nem csak a coiled coil/CSAH szakaszokon, hanem azokon kívül is. Néhány fehérje teljesen egyforma volt, akár különböző fajokon belül is. Az eredményeket később kibővítettem és az NCBI adatbázison is lefuttattam a hasonlósági keresést, illetve különböző hasonlósági mátrixokkal is próbáltam bővíteni a találatokat. Az eredmény az UniProton végzett 62-es BLOSUM mátrix esetében tűnt a legpontosabbnak, ezek mellé vagy számunkra jelentéktelen (coiled coil/CSAH szakasszal nem rendelkező) vagy még több teljes mértékben azonos találatokat kaptunk.

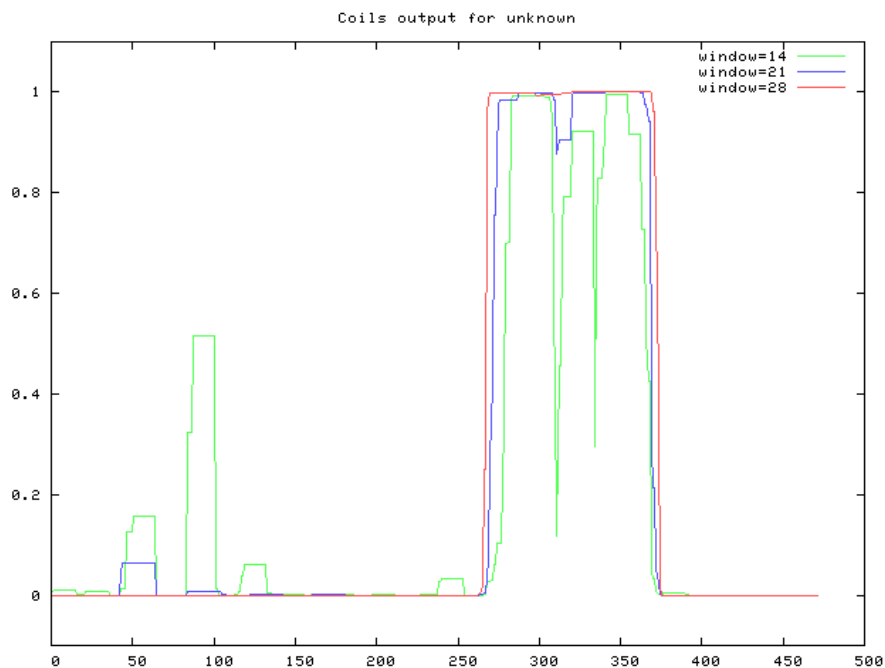
Különböző régió és domén felderítő algoritmusok segítségével megvizsgáltam, hogy a hasonlósági kereséskor találatul kapott fehérjék milyen térszerkezeti elemekkel és/vagy funkciókkal rendelkeznek.

A Coils-sal való coiled coil keresést többféle ablakolással végeztem el. Kereséskor egy valószínűségi értéket kapunk a szekvencia minden egyes eleméhez, hogy az mekkora eséllyel lehet része egy esetleges coiled coilnak. A különböző ablakolási méretek közül a legnagyobb (28-as) látszott a legmegbízhatóbbnak. Míg kisebb ablakolásnál több bizonytalan kiugrást is láthatunk, a szekvencia második felében minden esetben látható kiugrás, a különböző ablakméretek itt megerősítik egymást (2. ábra). A NONO_HUMAN esetében ezt az UniProt annotációja is megerősítette, bár ott megfigyelhető egy kis eltérés a jelzett szakasz hosszában.. A ProCoil és a LogiCoil segítségével megvizsgáltam hány szálú coiled coilal van dolgunk, ez mindenesetben dimernek (kétszálú) bizonyult. A fehérjét ezután RasMolban is ábrázoltam, a Protein Data Bank oldaláról letöltött pdb fájl alapján. A coiled coil szakaszt kiemeltem a SOCKET és a COILS szerver coiled coil predikciója alapján. (3. ábra)

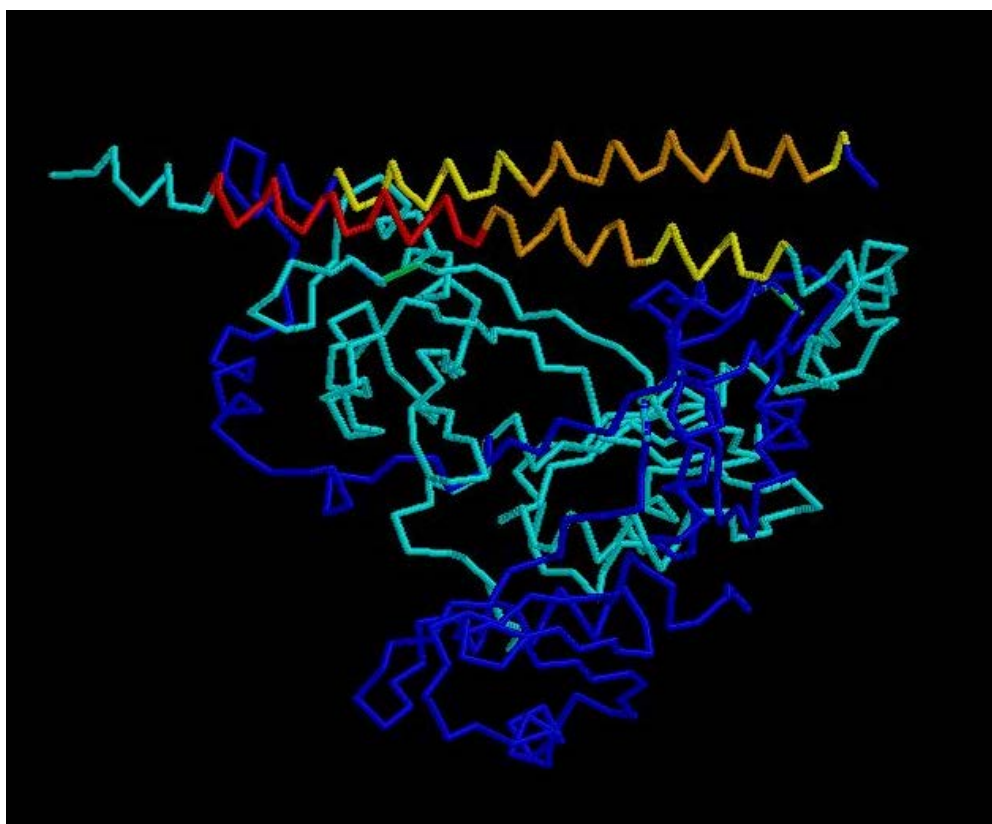
A CSAH keresés megerősítette az irodalmat, a jelenleg rendelkezésre álló algoritmusok minden coiled coil szakasz második felét CSAH-nek érzékelték. CSAH szakasz keresésnél a konszenzus szakaszt figyeltem, azaz azt a metszetet, ahol mindkét keresési módszer ezt a térszerkezetet prediktálta.

Rendezetlen szakaszokban nem ilyen egyértelmű a helyzet. A találatok itt is valószínűségeket adnak meg. Az IUPRED által javasolt 50%-os küszöböt figyelembe véve a fehérjék legnagyobb része rendezetlen, 70-90%-os esetben a szekvencia elején és végén látható rendezetlen szakasz, azonban ezek nem esnek egybe a coiled coil és CSAH szakaszokkal.

Az, hogy SFPQ (Splicing Factor, Proline- and Glutamine rich) heterodimerekről van szó magyarázza a konzerváltságot (saját készítésű cpp program alapján ez a két aminosav teszi ki a fehérje mintegy 20%-át, illetve a coiled coil szakasz 40%-át), mivel a teljes felszínen van kötőhely, azonban a CSAH-ot tartalmazó fehérjékre nem jellemző az ilyen szintű konzerváltság.



2. ábra: A Coils¹⁷ program grafikus kimenete a NONO_HUMAN esetén, a függőleges tengely a valószínűség, a vízszintes az aminosav helye a láncban. A 3 ablakolás (14/21/28) közül minden esetben a szekvencia második felében jelez egyértelműen coiled coilt. Kisebb ablakolás esetén látható néhány bizonytalan kiugrás



3. ábra: A NONO_HUMAN vizualizációja RasMol-ban (pdb azonosító: 3SDE¹⁵), kiemelve a SOCKET (sárga) és a COILS (piros) által prediktált Coiled Coil rész. A közös rész narancssárgával van jelölve. Az ábrán megfigyelhető még a ProCoil és a LogiCoil által is jósolt kétszálúság.

A különböző domének kereséséhez az SBASE-t használtam. Megfigyeltem, hogy a CSAH, coiled coil és rendezetlen szakaszok mellett a szekvenciák mindegyikén megjelenik az RRM (Rna Recognition Motif), amely többek között RNS kötő funkcióval rendelkezik. A friss szakirodalom rávilágított, hogy két RRM szakasz és egy coiled coil régió együtt a DBHS domént alkotják, melynek fontos szerepe van a transzkripció-regulációban, DNS-repairben, valamint számos biológiai folyamatot befolyásol, többek között a napi ritmust és a rák fejlődését.²⁵ A szekvenciákban található egyéb domének pontszáma alacsony volt, így ezekről nem lehet egyértelműen kijelenteni, megtalálhatóak a fehérjékben:

- Adeo Terminal: a replikálódó DNS végéhez kapcsolódó régió.
- Ribosomal 60S: A transláció elongáció szakaszában játszik szerepet.
- T-Complex Protein P11: a termékeny emlős herékben expresszálódik, a termékenyégben játszhat szerepet.
- Collagen triple helix: A szövetek kapcsolódási struktúrájában játszik szerepet.
- Bacterial inner membrane transport: többkomponensű rendszerek, amelyek szubsztrát-kötő fehérjéből, egy vagy két egymásra kölcsönösen homológ membrán fehérjéből és egy vagy két periférikus ATP-kötő fehérjéből állnak.
- Papillomavirus capsid L1: Papillomavírus, rákot és bőrképződményeket okozhat.
- For complement C1r/C1s, Uegf, Bmp1: Funkcionálisan sokszínű, többnyire fejlődésszabályozó proteinekben van jelen
- cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA: A phytochromeban jelenlevő domén, mely a fotoreceptorok egyik fajtája²³

Az teljes SwissProt adatbázison lefuttatott keresés alapján elmondható, hogy két RRM szekvencia megjelenése nem számít ritkának, azonban tekintettel, hogy a coiled coil szakaszok igen szegényesen vannak annotálva, egyelőre nem lehet további következtetéseket levonni a DBHS további megjelenéseiről.

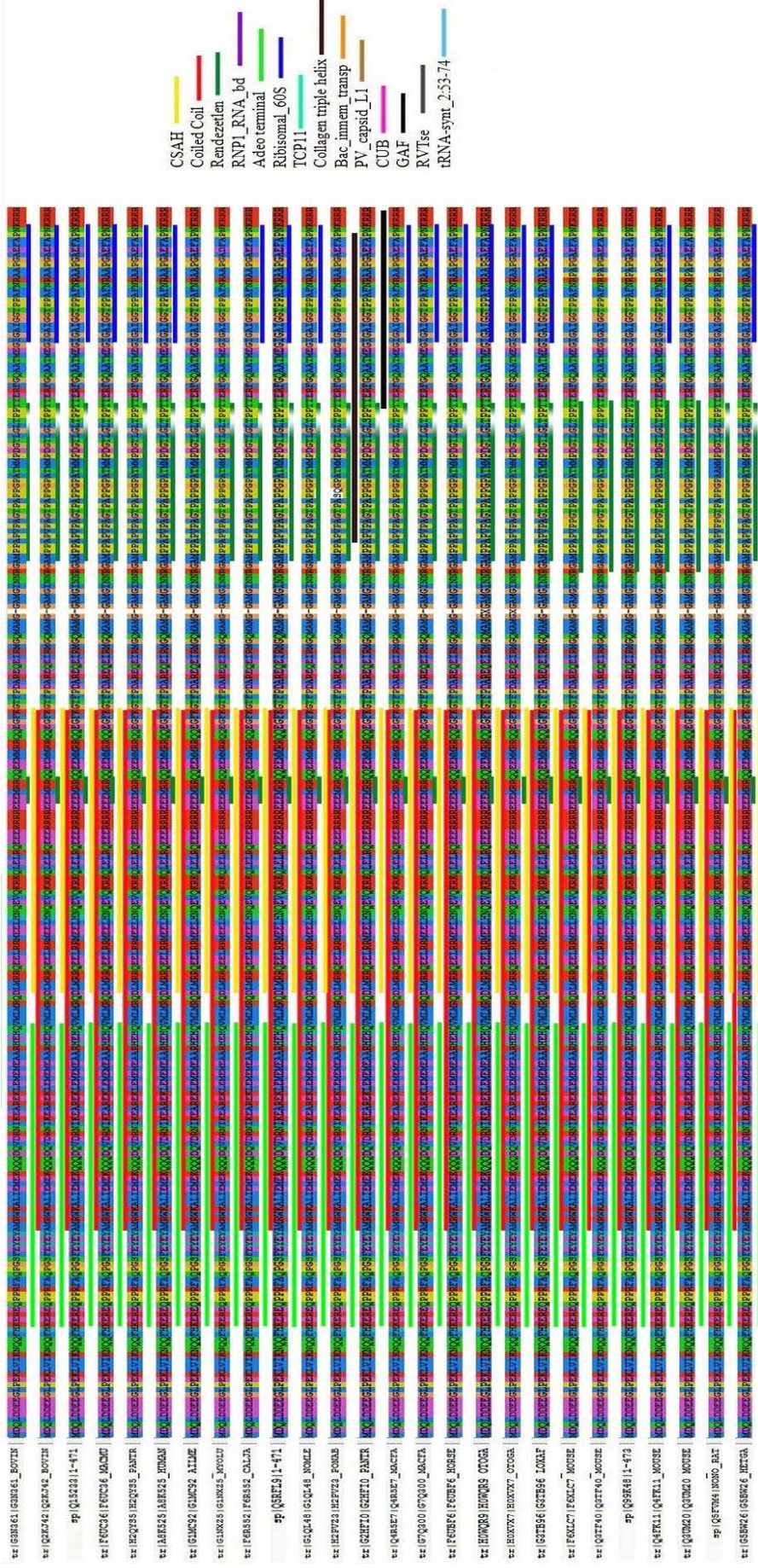
A filogenetikus fán (6. ábra) jól látszik, hogy a különböző fajokban fellelhető fehérjék viszonylag korán elágaznak. A NONO fehérjék a fejlődés egy korai szakaszában elváltak már, a patkányban és az egérben expresszálódók közelebb kerültek egymáshoz, az orángutánban és az emberben levő viszont kettő, már a fejlődés legelején elvált ágon vannak. Fontos megjegyezni, hogy a találatok jelentős része a TrEMBL-ből származik, ahol nem találhatunk részletes annotációt ezekről a fehérjékről, sem család, sem pedig ellátott funkciók szempontjából. A talált fehérjék nagymértékű hasonlósága miatt ez az eredmény kevésbé megbízható.

Vizsgáltam a szekvencia-konfliktus lehetőségét is. Azoknál a fehérjéknél, ahol fel volt jegyezve szekvencia konfliktus az összes programot még egyszer lefuttattam a módosításokkal. Az találatok nem változtak, így kijelenthető hogy ezek az anomáliák nem befolyásolják az eredményeket.

A találatokon többszörös illesztést hajtottam végre. Az képről (5. ábra) leolvasható, hogy a keresett motívumok egy erősen konzervált részt alkotnak. Az illesztésen ábrázoltam az összes talált domént és struktúrát. Azt a részt jelöltem coiled coilnak, ahol a valószínűség legalább 90% volt a legnagyobb ablaknál a Coils eredményei szerint. Az IUPRED által javasolt 50%-os küszöb esetén a fehérje legnagyobb része rendezetlennek bizonyult, 80-90%-os küszöb esetén a coiled coil/CSAH rész nem tekinthető rendezetlennek. A többszörös illesztésen ez utóbbit vettem ábrázoltam.

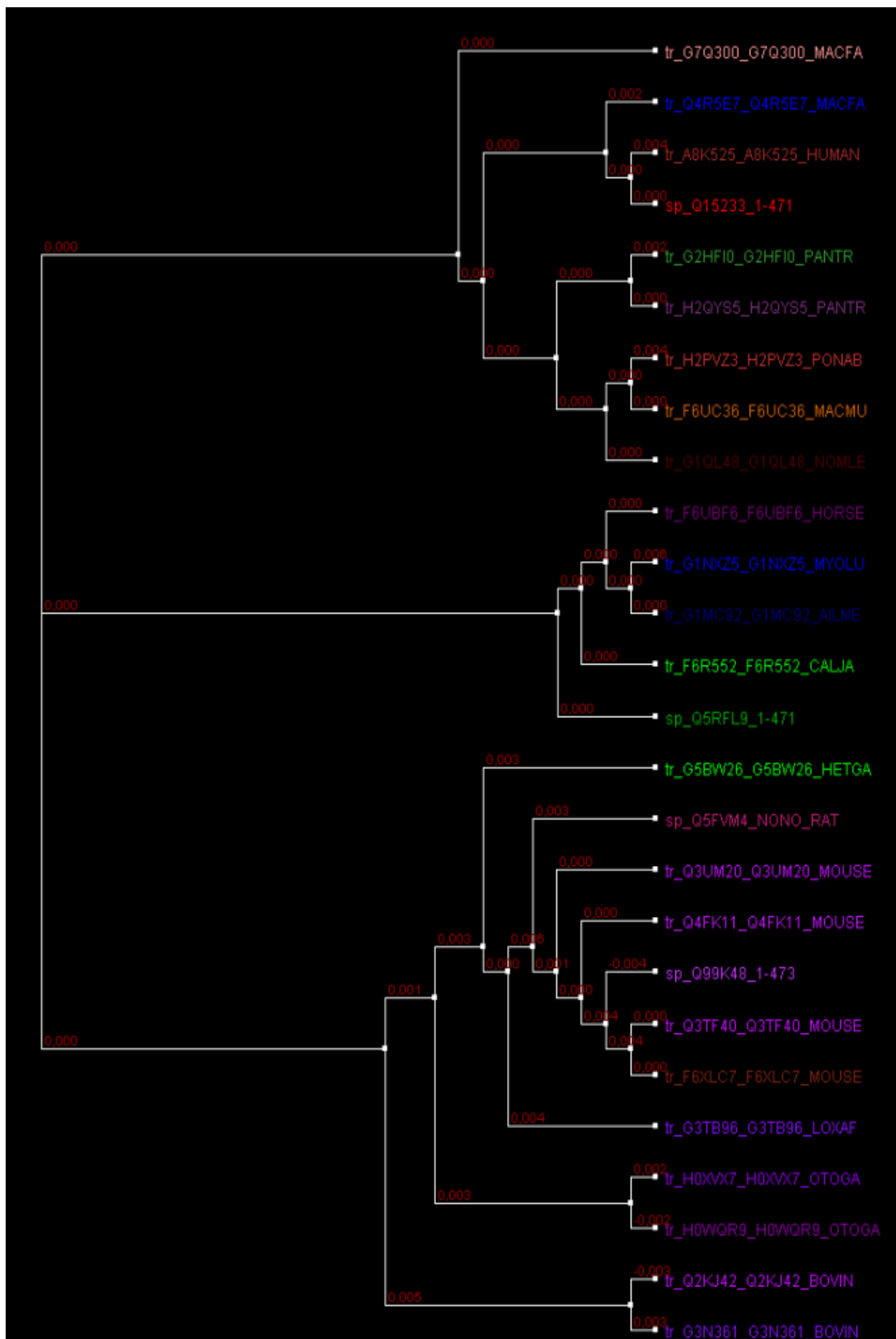
Az illesztés, a doménkeresés és a filogenetikus fa alapján megállapítható, hogy a nem SwissProt adatbázisból gyűjtött fehérjék közül 5 kivételével (G1NXZ5, G3N361, H0XVX7, F6XLC7, Q3TF40) jó eséllyel mindegyik a NONO családba tartozik, valószínűleg még nem került sor a feldolgozásukra. A kiemelt 5 fehérjéről sem lehet egyértelműen kijelenteni, hogy nem tartoznak ide, azonban közöttük szekvenciában és funkcióban is eltérés található szemben a többi fehérje relatív egységességével.

Az exon-intron határok keresésekor 4 féle mRNS variáns ált rendelkezésre az NCBI találatai alapján, amely (elvileg) képes kódolni a NONO_HUMAN-t. Megállapítható, hogy ebből 3 a NONO_HUMAN szekvenciánál egy jelentősen hosszabb szakaszt, míg 1 csak egy részét képes létrehozni (ábra). Az exon határok felszabdaltják a talált doméneket. A CSAH és az Adeo terminal eleje ugyan megegyezik egy exonhatárral, azonban a CSAH szekvenciát félbevágja egy intron. Az adeo terminal ezzel szemben végigkódolódik, ami megerősíti azt a várakozásunkat, hogy szerepet játszik a DNS replikációban, és mindig a kódoló rész határán helyezkedik el, ezért alacsony SBASE pontszáma ellenére lehet, hogy megjelenik a fehérjében.



4. ábra: A többszörös illesztés eredménye és a különböző domének elhelyezkedése a fehérjékben. Az ábra első fele a szekvenciák elejét, a második fele a szekvenciák végét mutatja. Megfigyelhető a coiled coil (piros) és a Charged Single Alfa-Helix átfedése. A coiled coil szakasz előtt minden esetben megjelenik két RNA Recognition Motif (lila), amelyek együtt a Drosophila Behaviour Human Splicing domént alkotják. Az ábra mellett található a többi domén/szerkezet színek kódja. Rövidítések:

CSAH: Charged Single Alfa-Helix	tRNA-synt_2: tRNA synthetase	PV capsid L1: PapillomaVirus capsid L1
RNP1_RNA_bd - RRM. RNA Recognition Motif	TCP11: T-Complex Protein	RVTse: Reverse Transcriptase
GAF: cGMP-specific phosphodiesterases, adenylyl cyclases abd FhlA	Bac inmem transp: Bacterial Inner Membrane Transport	
CUB: for Complement C1r/C1s, Uegf, Bmp1		



5. ábra: A filogenetikus fa. A különböző fajok más ágakra kerültek, így velük együtt a 4 korábban már annotált NONO fehérje is már a fejlődés korai szakaszában elvált egymástól. A TrEMBL (tr jelölés a fehérje ID előtt) adatbázisból származó fehérjék nagy valószínűséggel a NONO család eddig nem vizsgált tagjai. Fontos megjegyezni, hogy az erős konzerváltság miatt a filogenetika nem feltétlenül ad pontos eredményt.

8. Összefoglalás, kitekintés

A feladat során fehérjék másodlagos szerkezetében kerestem coiled coil, charged single alfa-helix és rendezetlen motívumokat. A kapott irodalom alapján választottam a NONO_HUMAN-t, mivel ezzel a fehérjével még nem végeztek sok vizsgálatot, viszont várhatóan tartalmaz coiled coil és CSAH szakaszokat. Különböző bioinformatikai módszerekkel kerestem vele homológ fehérjéket, vizsgáltam a szekvenciájukat (és ezen keresztül a struktúrájukat), funkciójukat illetve az evolúciós változásokat. Ez az első olyan munka, ahol egy teljes fehérjecsaládon végeztek bioinformatikai vizsgálatot a coiled coil/CSAH/rendezetlen szakaszok vizsgálata céljából.

A munka megerősítette a szakirodalomban olvasottakat: az egyes konformációk könnyen egymásba hajlanak, nehezen megkülönböztethetők, nincs egységes algoritmus annak megállapítására, hogy milyen szerkezetet látunk. A homológ szakaszok keresése során többféle beállítást is megpróbáltam: a túl erős feltételeket állító algoritmusok elvesznek jó eredményeket, ellenkező esetben túl sok hibás vagy nem releváns szekvenciát kaptam. Az adatbázisok töredezettségét mutatja, hogy a keresés során olyan fehérjéket is találtam, amelyeket még nem annotáltak. A hasonlósági keresés eredményei tartalmaztak coiled coil régiókat, CSAH szekvenciákat és rendezetlen szakaszokat. Ezeknek a fehérjéknek jelentős része valószínűleg a NONO családba tartozik, csak korábban még nem végeztek rajtuk részletes vizsgálatokat.

A többszörös illesztés eredményéből jól látható, hogy a vizsgált fehérjék coiled coil régiói nagyon kevés különbséget mutatnak szekvenciájukat illetően, illetve az aminosavak eloszlásának mintázata is csaknem minden esetben megegyező. Mindez arra utal, hogy adott molekuláris funkció betöltéséhez evolúciósan meglehetősen konzervált rész szükséges a vizsgált fehérjék esetén.

A kísérleteket érdemes továbbfolytatni, más módszereket kipróbálni. Az NCBI adatbázisra történő kiterjesztés nem feltétlenül járt sikerrel, azonban érdekes lehet PSI-BLAST módszerrel keresni homológ szekvenciapárokat. Ez a BLAST összerendezés-kereső program kiterjesztése, mely a többszörös illesztés alapján létrehozott profilt használja kereséshez. Szintén meg lehet próbálni a kiterjesztést Hidden Markov Model segítségével, bár az ilyen jellegű analízishez egyelőre kevés adat áll rendelkezésre. Fontos lehet a DBHS a coiled coil és a CSAH szakaszok közötti összefüggést tovább vizsgálni. Érdemes lehet más fehérjecsaládot is bevonni a vizsgálatba.

A coiled coil-ok fontos szerkezeti motívumok, amelyek számtalan fontos kölcsönhatásban szerepet játszanak, amelyeket biokémiailag, terápiásan és a nanotechnológiában is alkalmazni lehet. A Charged Single Alfa-Helix-ek pontos funkcióit még nem ismerjük, viszont mivel az emberben gyakran előforduló, fontos funkciókat ellátó fehérjékben jelenik meg, érdemes foglalkozni velük. Elsőként azonban jobban meg kell értenünk azokat az interakciókat, amelyek a coiled coil bizonyos átalakulásaihoz vezetnek, amelyhez további kísérletek és bioinformatikai ötletek szükségesek.

9. Irodalomjegyzék

- ¹“Non-POU domain-containing octamer-binding protein”, <http://www.genecards.org/cgi-bin/carddisp.pl?gene=NONO&search=NONO>, Oct. 31, 2012 [May. 1, 2012].
- ² Dániel Süveges, Gábor Tóth, László Nyitray, Zoltán Gáspári. „Charged single alfa-helix: A versatile protein structural motif”. *Proteins*, vol. 74, pp. 905-916, 2009.
- ³ Y.Bruce Yu. "Coiled-coils: stability, specificity, and drug delivery potential". *Adv Drug Deliv Rev*, vol. 54(8), pp 1113-1129, 2002.
- ⁴Balázs Szappanos, Dániel Süveges, László Nyitray, András Perczel, Zoltán Gáspári. „Folded-unfolded cross-prediction and protein evolution: The case study of coiled-coils”. *Febs Letters*, vol. 584, pp. 1623-1627, 2010.
- ⁵Zoltán Gáspári, Dániel Süveges, András Perczel, László Nyitray, Gábor Tóth. „Charged single alpha-helices in proteomes revealed by a consensus prediction approach”. *Biochimica et Biophysica Acta*, vol. 1824(4), pp. 637-646, 2012.
- ⁶Misha Soskine, Dan S. Tawfik. „Mutational effects and the evolution of new protein functions”. *Nature Reviews Genetics*, vol. 11, pp. 572-582, 2010.
- ⁷Zoltán Gáspári, László Nyitray. „Coiled coils as possible models of protein structure evolution”. *BioMol Concepts*, vol. 2, pp. 199-210, 2011.
- ⁸Jody M. Mason, Mark A. Schmitz, Kristian M. Müller, and Katja M. Arndt: "bCIPA: Coiled Coil input form". <http://www.molbiotech.uni-freiburg.de/bCIPA/> Nov. 14, 2012 [2010]
- ⁹“Uniprot”, <http://www.uniprot.org/> , Oct. 31, 2012 [2012].
- ¹⁰Nick V. Grishin. „Fold change in evolution of protein structures”. *Journal of structural biology*, vol. 134, pp. 167-185, 2001.
- ¹¹“Non-POU domain containing, octamer-binding [Homo Sapiens] - Gene -NCBI”, http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=retrieve&list_uids=4841, Oct. 18, 2012[Nov. 2, 2012].
- ¹²“Non-POU domain-containing octamer-binding protein”, <http://www.uniprot.org/uniprot/Q15233>, Oct. 31, 2012[Nov. 2, 2012].

- ¹³"ClustalW and ClustalX Multiple Sequence Alignment" <http://www.clustal.org/clustal2>, Oct. 15, 2012 [2010]
- ¹⁴"Phylogeny.fr: Home", <http://www.phylogeny.fr/>, Nov 1, 2012 [2003]
- ¹⁵"RCSB Protein Data Bank- RCSB PDB -3SDE Strucutre Summary"
<http://www.rcsb.org/pdb/explore/explore.do?structureId=3SDE> Nov. 15, 2012 [Nov. 13, 2012]
- ¹⁶Tom Alber, Derek N. Woolfson. Predicting oligomerization states of coiled coils". *Protein Science*, vol. 4, pp. 1596-1607, 1995.
- ¹⁷Lupas, A., Van Dyke, M., and Stock, J.: "COILS Server",
http://www.ch.embnet.org/software/COILS_form.html, Nov 1, 2012
- ¹⁸"The R Project for Statistical Computing", <http://www.r-project.org/>, Nov 1, 2012 [2012]
- ¹⁹Derek N. Woolfson, Thomas L.Vincent, Peter J. Green „LOGICOIL - Multi-state prediction of coiled-coil oligomeric state". *Bioinformatics online*, Nov 4, 2012.
- ²⁰John Walshaw: "Socket, at the Woolfson Group, University of Bristol"
<http://coiledcoils.chm.bris.ac.uk/socket/server.html>, Nov. 15, 2012
- ²¹Zoltán Gáspári: "CSAH Server", <http://himalia.chem.elte.hu/cgi-bin/csahserver.cgi>, Nov 1, 2012 [2009]
- ²²Zsuzsanna Dosztányi, Veronika Csizmók, Péter Tompa and István Simon: "IUPRED",
<http://iupred.enzim.hu/>, Nov 1, 2012
- ²³Mircea Pacurar, Laszlo Kajan, Sandor Pongor: "SBASE Web", <http://hydra.icgeb.trieste.it/sbase/> , Nov 1, 2012 [2006]
- ²⁴"Bioinformatics Tools for Sequence Translation", <http://www.ebi.ac.uk/Tools/st/>, Nov 1, 2012 [2012]
- ²⁵Daniel M. Passon, Mihwa Lee, Oliver Rackham, Will A. Stanley, Agata Sadowska, Aleksandra Filipovska, Archa H. Fox, Charles S. Bond. „Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation". *PNAS*, vol 109 no. 13, pp. 4846-4850, 2012

10. Köszönetnyilvánítás

Köszönettel tartozom konzulensemnek, Dr. Gáspári Zoltánnak a kapott útmutatásáért, a felmerülő nehézségek esetében nyújtott segítségéért és javaslataiért és hogy az elmúlt évben felügyelte a munkámat.

"Nyilatkozat a pályázó egyéni hozzájárulásáról a pályaműben leírt eredményekhez"

Alulírott Dobson László a dolgozatban leírt munkafolyamatok közül a:

-BLAST analízis

-Többszörös illesztés

-Filogenetikus fa készítése

-Különböző régió és doménfelderítő algoritmusok futtatása (COILS, COILS eredmények kiértékelésére írt cpp kód, LogiCoil, PairCoil, SOCKET, SCAN4CSAH, FT_CHARGE, IUPRED, IUPRED kiértékelésére írt cpp kód, SBASE, különböző szempontok szerint a teljes SwissProt adatbázison való keresésre írt python kód)

-Exon- és intron határok keresése

-In Silico fordítás az EMBOSS programcsomaggal

-Aminosaveloszlás vizsgálatára írt cpp kód

végeztem el.

.....

aláírás

Budapest, 2013 január 08.