

XXXI. OTDK Biológia Szekció

Az amiloidképző fehérjeszakaszok részletes bioinformatikai vizsgálata

TDK dolgozat, 2012

Hatos András, Molekuláris Bionika BSc hallgató



Pázmány Péter Katolikus Egyetem, Információs Technológiai Kar

Belső Konzulens: Gáspári Zoltán



*Magyar Tudományos Akadémia Természettudományi Kutatóközpont
Enzimológiai Intézet*

Témavezető: Kalmár Lajos

Tartalom

1.	Bevezetés	3
1.1.	Az amiloidózis általános áttekintése	3
1.2.	Szisztémás amiloidózis	7
1.3.	Szervi lokalizációjú amiloidózis	8
1.4.	Fehérje misfolding és sejten belüli amiloidózis	9
1.5.	Az amiloid képződés szerkezeti alapjai, bioinformatikai megközelíthetősége	10
2.	Célkitűzések	12
3.	Módszerek	13
3.1.	Adatgyűjtés	13
3.2.	Amiloidogén szakaszok feldolgozása	18
3.2.1.	Aminosav kompozíció elemzése	18
3.2.2.	Aminosav index szerinti kategorizálás	18
3.2.3.	Aminosav index értékek statisztikai szűrése	20
4.	Eredmények	22
4.1.	Aminosav kompozíció	22
4.2.	AAindex eloszlások vizuális kiértékelése, előszűrés	24
4.3.	AAindex eloszlások statisztikai feldolgozása	25
5.	Diszkusszió	27
6.	Köszönetnyilvánítás	28
7.	Melléklet	29
8.	Irodalomjegyzék	34
9.	Nyilatkozat	36

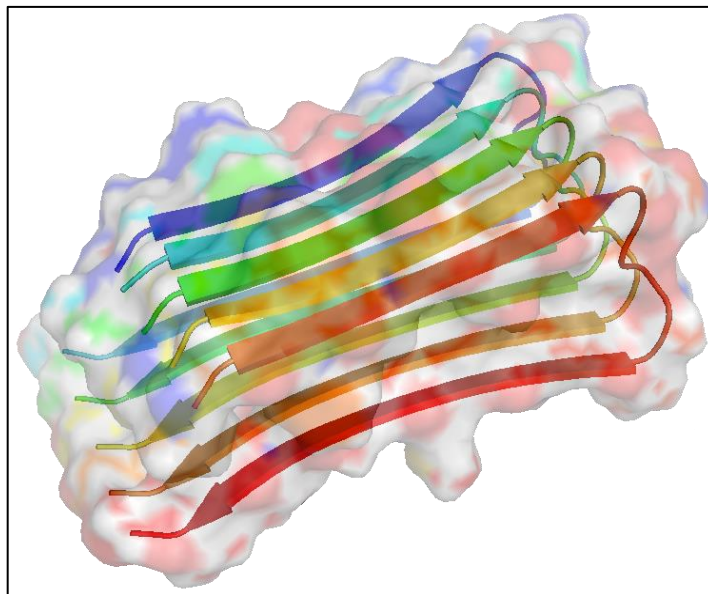
1. Bevezetés

1.1. Az amiloidózis általános áttekintése

Az elmúlt évtizedben számos klinikailag különböző betegséget kötött össze a közös patomechanizmusuk, ezen betegségeken bizonyos fehérjék kóros feltekeredés (misfolding) miatt speciális aggregátumokat, ún. amiloidot képeznek, ami mai tudásunk szerint toxikus a sejtre nézve [1]. Ezekre a betegségekre együttesen ezért az amiloidózis kifejezést használjuk, mert bár a kialakításában szerepet játszó fehérjék és a klinikai tünetek igen sokfélék lehetnek, a sejten belüli változások mechanizmusa azonos. Amiloidózissal alakul ki több igen gyakori betegség: számos neurodegeneratív betegség (pl. Alzheimer- és Parkinson-kór), bizonyos diabetes formák, a Huntington-kór és a vese-dialízishez köthető szisztémás amiloidózis is [2-4]. A közös patomechanizmussal járó betegségek tehát igen gyakoriak, orvosbiológiai jelentőségük nagy, kialakulásuk megértése legalább annyira fontos feladata a tudománynak, mint a betegségek elleni gyógymód megtalálása.

Jelenlegi tudásunk szerint az amiloidózis hátterében elsősorban olyan fehérjék állnak, melyek valamilyen környezeti hatásra a normál szerkezetüktől eltérő, kóros szerkezetet vesznek fel, mely állapotban lehetőségük van polimerek, ún. amiloid szálak kialakítására [5]. Az eddig ismert amiloidokban a stabil, oldószereknek, hőhatásnak is igen ellenálló fibrilláris szerkezetet egymás mellé rendeződő béta-redő struktúrák alkotják (1. ábra). Az amiloidot képző fehérjék natív szerkezete általában kevésbé stabil, mint az átlagos globuláris fehérjéké, gyakran találunk köztük rendezetlen fehérjéket. A rendezetlen fehérjék alapvető tulajdonsága, hogy nem rendelkeznek jól definiálható harmadlagos szerkezettel, sok különböző szerkezeti állapot között fluktuálnak. Ez a tulajdonságuk kulcs faktor lehet az amiloid képződésben, hiszen míg egy jól definiált harmadlagos szerkezettel rendelkező globuláris fehérjének jelentős energiatöbblet szükséges a stabil szerkezetből való kimozdításához, és a kóros konformáció felvételére, addig a rendezetlen fehérje konformációs sokaságából több olyan állapot is lehet, mely nagyon könnyen épülhet be az amiloid szálba. A rendezetlen fehérjéket e tulajdonságuk miatt több

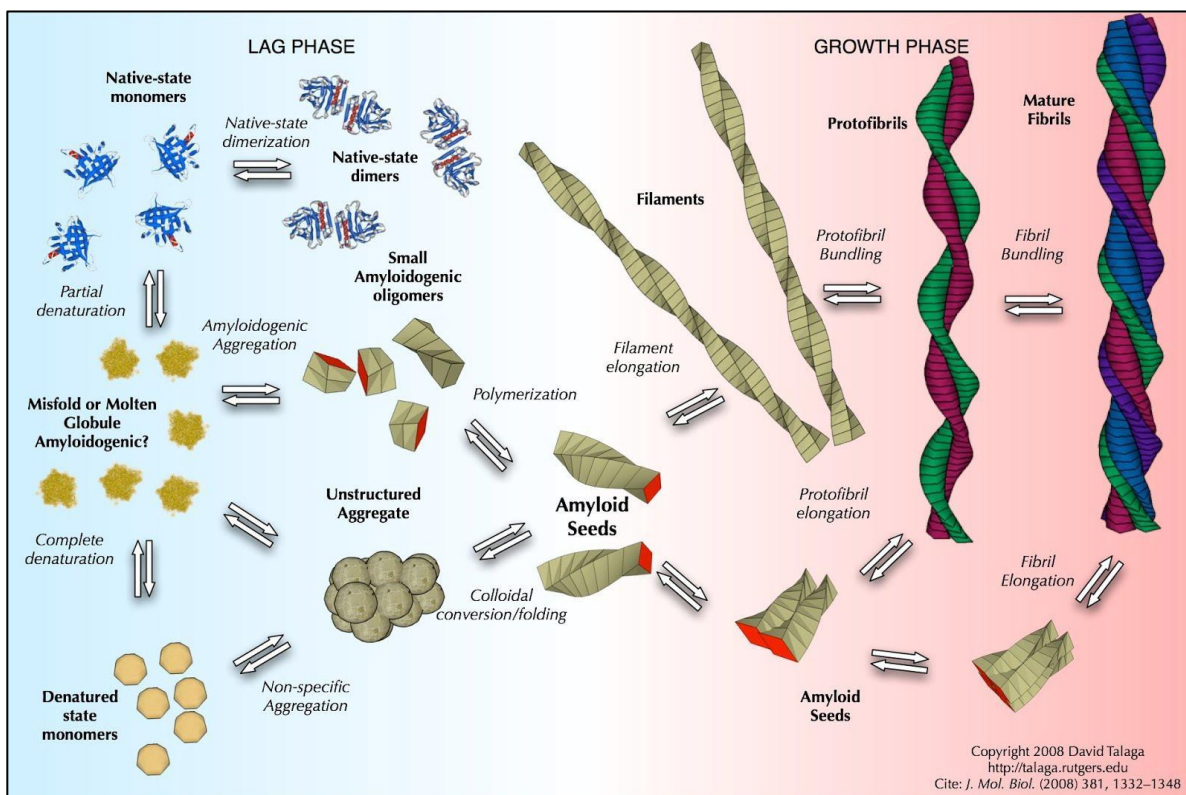
tanulmány is a sejtre nézve veszélyes fehérjének tartja, azonban fontos leszögezni, hogy az ismert rendezetlen fehérjének csak töredéke képes valóban amiloid képzésre.



1. ábra. Az amiloid szálak harmadlagos szerkezetének példája. A szál felépítő azonos fehérjék egymás mellé rendeződve, paralell vagy antiparalell béta-lemez szerkezetet képeznek. Az ábrán a különböző színek külön fehérje láncokat jelentenek, a fehérje amiloid szálba nem beépülő részei nem lettek ábrázolva. Forrás: internet.

Az amiloid képződés kulcs mozzanata az első néhány aggregátum, amiloid szál képződése. Az *in vitro* kísérletekben ezeket az oligomereket seedeknek nevezzük, melyek az amiloid képződés kísérletes tanulmányozásakor nagymértékben elősegítik a szálképződés megindulását. Úgy gondoljuk, hogy az élő sejten belül is egy fontos lépés lehet a kritikus mennyiségű amiloid szál megjelenése, egy szinten még a sejt javító rendszerei (elsősorban a chaperon fehérjék) képesek eliminálni a képződött kóros aggregátumokat, de ennek telítődése után az amiloid képződés nagy valószínűséggel elszabadul, kikerül a sejt kontrollja alól, és exponenciális növekedésbe kezd [6]. A képződött amiloid szálaknak elsősorban a végei azok, melyek a további monomerek beépülését elősegítik. Ez a kísérletes amiloid képződésben is kimutatható, az amiloid szálak növekedése sokkal gyorsabban elindul és nagyobb mértékű, ha nem hosszú, intakt amiloid szálakat használunk, hanem ezeket kisebb darabra törve (pl. mikrohullám segítségével) alkalmazzuk seedként. A kísérletes úton és a sejten belül kialakult amiloid szálak is igen stabilak,

ellenállóak, valószínűleg ez az, ami miatt a sejtekre is toxikussá válnak, hiszen a felgyülemelő aggregátumot a sejtek már sem lebontani, sem elkülöníteni nem tudják [7]. Az amiloid képződés folyamatát a 2. ábra foglalja össze.

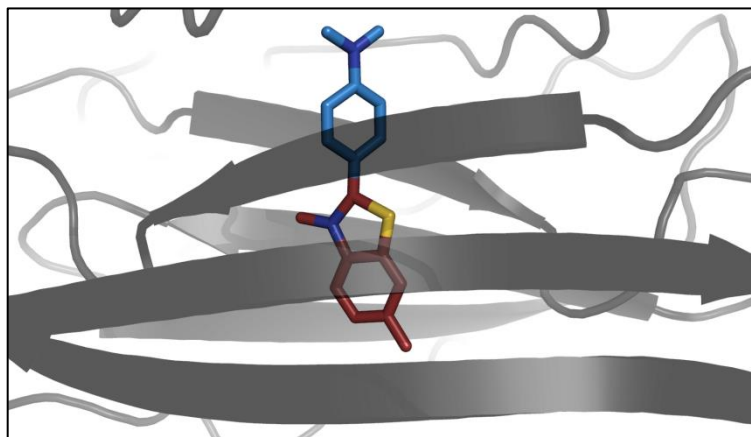


2. ábra. Az amiloid képződés folyamatának sematikus vázlata. Az amiloid képződés alapvetően két fázisra osztható, az első "lag" fázisban az amiloid képzésre hajlamos fehérjék kóros feltekeredése történik meg, melynek következtében oligomerek, amiloidogén kis polimerek képződnek. A megfelelő hosszúságú amiloid szálak (seedek) indítják be a második, növekedési fázist. Ebben a már kialakult amiloid szálakhoz épülnek hozzá a monomer vagy oligomer egységek, és növesztik a szál hosszát. A hosszú fibrilláris szerkezetek egymás mellé rendeződésével alakul ki a végső, igen stabil amiloid szál forma. Az ábrán található kifejezéseket angol nyelven hagytam, mert a legtöbb kifejezésre még nincs igazán magyar megfelelő [8].

Ahogy az a 2. ábrán is jól látható, az amiloid képződés két fázisra osztható. A két folyamat *in vitro* körülmények között eltérő sebességgel zajlik. A nagy koncentrációjú amiloidogén fehérje oldatban órák vagy akár napok szükségesek az első fázis kialakulásához, vagyis a kritikus mennyiségű seed keletkezéséhez, amely után a második fázisban történő szálképződés már jóval gyorsabb folyamat [9]. Természetesen élő sejtben a folyamat kissé más módon zajlik, de

úgy tűnik, hogy in vivo is a folyamat első szakasza az idő meghatározó lépés. Az élő szervezetben kialakuló amiloidózis egyik fontos befolyásoló tényezője szintén az adott fehérje koncentrációja, de a betegségek kialakulásának ideje függhet az amiloidózis első szakaszának sebességétől is. Azok az örökletes mutációk, melyek a Parkinson-kór korai kialakulásához vezetnek (50-60 éves kor helyett akár 20-30 éves korban is kialakulhat), in vitro körülmények között is hamarabb átesnek az amiloidózis első, azaz "lag" fázisán, és hamarabb indul el a vizsgált mintában a szálképződés [10]. Ebből arra következtethetünk, hogy az élő sejtekben is az első fázis a kritikus az amiloidózishoz köthető betegségek kialakulásában. Ez az a lépés ami sokszor akár évekig, évtizedekig is eltarthat, és amennyiben a kritikus mennyiségű seed felgyülemlett a sejtben, akkor az már nem tud megszabadulni tőle, és elindul a pusztulás útján.

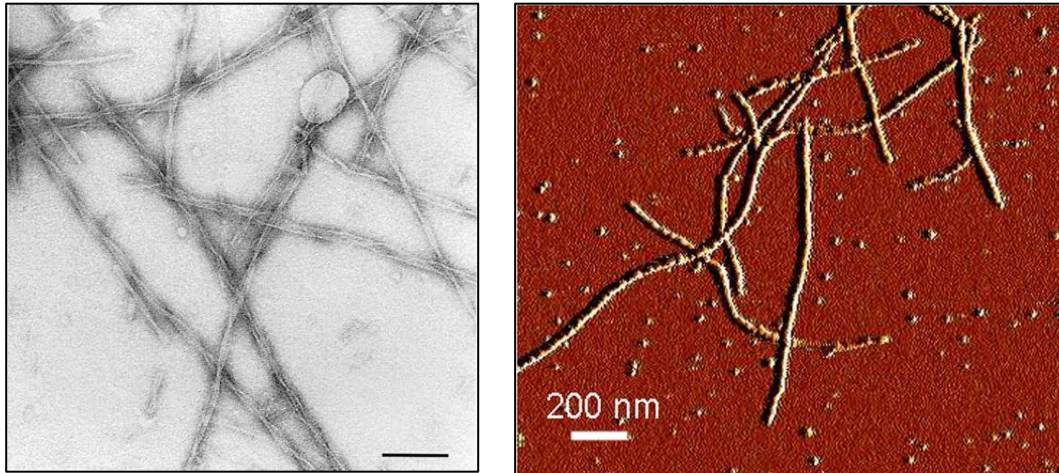
Az amiloid növekedésének in vitro követésére a thioflavin t (THT) tesztet használják. Ez a heterociklusos kis molekula a merev béta-lemez struktúrákhoz köt, így az amiloid képződést időben nyomon követve kvantitatív méréseket is lehet végezni (3. ábra).



3. ábra. A THT kötődése az amiloid aggregátumhoz. Az amiloid kimutatásban használt THT teszt során a thioflavin t molekula a béta lemezes struktúrába épül, mely által nagymértékben csökken a molekula mozgása. A kézzel jelölt dimethylaminobenzyl rész nem tudja excitációs töltését átadni a pirossal jelölt benzothiazole résznek, emiatt tapasztalható a molekula fluoreszcens jelében egy vörös eltolódás, mely kísérletesen jól mérhető. Forrás: internet.

Mivel a THT molekula egyformán köti a rövid és a hosszú amiloid szálakat, így segítségével csak a képződött amiloid mennyisége mérhető, az amiloid szálak morfológiája nem határozható

meg. Miután egyes betegségekben az amiloid szálak felépítésének is tulajdonítanak jelentőséget, vizsgálati módszerek alakultak a morfológia vizsgálatra is. A néhány nanométer átmérőjű, de sokszor akár mikrométeres hosszúságú amiloid szálak makroszerkezetének vizsgálatára elektronmikroszkópos (THT vagy negatív ozmium-tertoxid festés) vagy atomerő mikroszkópos vizsgálatokat alkalmaznak (4. ábra).



4. ábra. Amiloid szálak morfológiájának vizsgálata. Bal oldalon negatív ozmium-tetroxiddal festett elektronmikroszkópos kép látható (a mérték 100 nm hosszú), jobb oldalon pedig a fibrillumok atomerő-mikroszkópos képe. Forrás: internet.

Az amiloidózisokat a szervezetben való megjelenési helyük alapján csoportokba oszthatjuk, így beszélhetünk szisztémás (teljes szervezetre kiterjedő), szervi lokalizációjú és specifikus sejteken belül kialakuló amiloidózisról.

1.2. Szisztémás amiloidózis

A szervezet egészére kiterjedő amiloidózisok alapja, hogy az extracelluláris térben valamilyen amiloidogén fehérjének felborul az egyensúlya, és a megnövekedett mennyiség következtében amiloid plakk lerakódások keletkeznek a test több pontján [11]. A teljes szervezetre kiterjedő amiloidózisok egyik leggyakoribb változata a béta-2-mikroglobulinhoz kötött szisztémás amiloidózis. Ez a fehérje a minden sejtünkön kifejeződő MHC I-es osztályú fehérjekomplex része, ezáltal igen fontos eleme az immunrendszer saját / nem saját elkülönítő rendszerének. A

fehérje egy transzmembrán komplex extracelluláris régiójának része, melyről a fehérje könnyedén disszociálhat. Ilyenkor a béta-2-mikroglobulin a vérbe kerül, ahonnan a vese kiválasztó funkciója távolítja el. Ennek a természetes folyamatnak köszönhetően egészséges emberekben a fehérje koncentrációja az extracelluláris térben és a vérkeringésben igen alacsony. Vesebetegek esetében azonban a fehérje kiválasztása sérül, a súlyos veseelégtelenségben szenvedők hemodialízis (művese) kezelésében pedig teljes mértékben megszűnik. A jelenlegi művese technológiák képesek megszabadítani a vért minden káros kismolekulától, bizonyos ionoktól, azonban a fehérjeként alapvetően kis méretű, de az ionokhoz, bomlástermékekhez képest mégis óriási béta-2-mikroglobulin fehérjét nem képesek kiszűrni. Emiatt az akár éveken át tartó művese kezelés a fehérje szervezeten belüli felgyülemeléséhez vezet, az extracelluláris térben a fehérje koncentráció eléri azt a kritikus koncentrációt, melynél az amiloid képződés beindulhat, és megkezdődik a plakkok lerakódása. A béta-2-mikroglobulinhoz köthető szisztémás amiloidózis igen sok szervre, szövetre kiterjed, elsődlegesen érintettek az érfalak, az ízületek és a szívburok.

A késői immunválasz következtében keletkező antigén-antitest komplexek felgyülemelése is okozhat szisztémás amiloidózist. Az akut fertőzések következtében keletkező immunkomplexek eliminálásával a szervezet könnyedén megküzd, azonban egy sok éven át tartó, elhúzódó krónikus betegségben (pl. gyulladásos betegség) keletkező immunkomplex mennyiség telíti a szervezet kiürítő rendszerét, és ilyenkor megindul az immunkomplexek képezte amiloid plakkok lerakódása különböző szövetekben.

A szisztémás amiloidózisok főként a fizikai tér elfoglalásával, más szövetek összenyomásával, azok megfelelő keringésének zavarásával okoznak egészségügyi problémákat. A lerakódások következtében elsősorban szív-érrendszeri és mozgásszervi betegségek alakulnak ki, melyek tovább súlyosbítják az egyébként a szisztémás amiloidózist kiváltó betegségben szenvedő paciens állapotát.

1.3. Szervi lokalizációjú amiloidózis

Az amiloidózisok egy szervre lokalizálódó fajtáinak kialakulása nagyjából megegyezik a szisztémás amiloidózis kialakulásával. Mindkét típusban valamilyen amiloidogén fehérje

extracelluláris térben való felgyülemzése váltja ki az amiloidózist. Az egy bizonyos szervre lokalizálódó amiloidózis esetében a fehérje termelődése az érintett szervben nagyon nagymértékben nő és a fehérje nagyon kis mértékben távozik a szervből a vérkeringés által. Így a kritikus koncentráció csak a szervben belül, annak extracelluláris állományában lép fel, és az amiloid plakkok lerakódása arra a szervre korlátozódik. A kiváltó okok is hasonlóak, mint szisztémás amiloidózisnál, legtöbbször elhúzódó, krónikus betegségekhez kapcsolt (gyulladás, daganatos megbetegedés) túlzott fehérje termeléssel, felgyülemléssel kapcsolatos. Ebbe a csoportba tartozik a kalcitonin lerakódás miatt kialakuló pajzsmirigy-elégtelenség, és az amylin hasnyálmirigyben történő lerakódása miatt kialakuló II-es típusú cukorbetegség is.

1.4. Fehérje misfolding és sejten belüli amiloidózis

Az elmúlt évtized felismerései nyitották meg a tudomány kapuit számos betegség, elsősorban az idős korral együtt járó neurodegeneratív betegségek megértésében. Több, akár évszázadok óta ismert, klinikailag jól karakterizált kórképpel kapcsolatban derült fény a hasonló patomechanizmus, az intracelluláris amiloidózis kóroki meglétére. Ezen betegségek hátterében a sejten belüli kóros, nem megfelelő fehérje feltekeredés, az ún. “misfolding” jelensége áll. Valószínűleg a szisztémás és szervi amiloidózisokhoz hasonlóan a sejten belüli amiloid képződésben is szerepe van az amiloidogén fehérje túltermelésének, azonban ennél a formánál az egyéb faktorokat, elsősorban környezeti hatásokat tesznek felelőssé. Szintén komoly szerepet tulajdonítanak a sejt felborult homeosztázisának, elsősorban a csökkent hatékonyságú javító és elimináló rendszereknek (chaperon rendszer, proteasóma rendszer) a sejten belüli amiloidózis kialakulásában.

A környezeti hatásra kialakuló intracelluláris amiloidózis legjobb példái a prion fehérje által kiváltott neurodegeneratív betegségek, melyet emberben Creutzfeldt–Jakob szindrómának, vagy Kuru-kuru betegségnek nevezünk, állatokban pedig különböző szivacsos agyvelő gyulladásokat okoz (pl. bovine spongiform encephalopathia, BSE) [2]. Ebben a betegségben a központi idegrendszerben található egyelőre még nem teljesen ismert funkciójú prion fehérje kóros feltekeredése okozza a betegség kialakulását. A vad típusú fehérje magától nem képez amiloidot, a folyamat elindításához szükség van az amiloid valamilyen forrásból történő

bevitelére, amely kóros szerkezeti változat (vagyis gyakorlatilag a korábban már említett amiloid seed). A normál szerkezetű és szekvenciájú fehérjét használva építőanyagként növeszti az amiloid szálat [12]. A leggyakrabban neurodegeneratív betegséghez vezető más intracelluláris amiloidózisoknál a betegséget kiváltó tényező nem ennyire egyértelmű [13]. Vannak esetek, amikor egy-egy ismert örökletes mutáció az eleve amiloidogén fehérjében elősegíti a kóros konformáció kialakulását, ezzel nagymértékben meggyorsítva a betegség kialakulását. Ezt figyelhetjük meg a Parkinson-kór kialakulásában is. Ebben a betegségben főként az alfa-synuclein fehérjét teszik felelőssé a központi idegrendszer bizonyos sejtjeiben lerakódó amiloid plakkok (szövettani megnevezésében Lewy-testek) kialakulásáért. A betegség megjelenése idős korban jellemző, elsősorban a mozgás koordinációért felelős agyi struktúrák degenerációja történik meg, ennek megfelelően a klinikai képre is a mozgás, beszéd érintettsége jellemző. Az időskori Parkinson-kór kialakulásáért sok környezeti hatást próbáltak már felelőssé tenni, melyek között szerepelnek nehézfém szennyezések, fizikai behatások (fejsérülések) is. Az esetek bizonyos részében azonban (kb. 15%) a Parkinson-kór jóval hamarabb jelentkezik, akár már fiatal vagy középkorban. Ilyenkor szinte minden esetben lehet találni olyan örökletes mutációt az alfa-synuclein fehérjében (ilyen ismert mutációk: A53T, A30P, E46K), melyekről kimutatható, hogy *in vitro* körülmények között is elősegítik az amiloid képződést. Ez esetben tehát nem annyira a környezeti tényezők, mint inkább egy genetikai faktor az, ami kiváltja a betegség, az amiloid plakkok keletkezését.

1.5. Az amiloid képződés szerkezeti alapjai, bioinformatikai megközelíthetősége

Ahogy azt már a korábbiakban is említettem, az amiloid szálak paralell vagy antiparalell béta-redő szerkezetből felépülő béta-lemezes szerkezetűek. A béta-redő és a béta-lemez szerkezet is igen elterjedt szerkezeti egység a harmadlagos szerkezettel rendelkező globuláris fehérjékben, azok nagy része mégsem amiloidogén. Ennek elsősorban az az oka, hogy a globuláris fehérjék a helyes szerkezetük felvételével olyan alacsony energiaállapotba kerülnek, melyből nagyon nehéz kizökkenteni őket, így nincs lehetőség arra, hogy kóros feltekeredéssel eltérő, amiloid képző szerkezetet vegyenek fel. Az amiloid képző fehérjékben a bennük található amiloidogén

fehérje szakaszok a legtöbbször rendezetlen vagy metastabil régiókban találhatók. Ezek a régiók eleve csak nagyon kis mértékben vagy egyáltalán nem vesznek fel harmadlagos szerkezetet, nincsenek alacsony, stabil energiaállapotban. Természetesen mivel a sejt számára toxikus az amiloid képződés, számos mechanizmus alakult ki ezen szakaszok kordában tartására, ennek kulcsfontosságú elemei a chaperon rendszerek, a helytelen feltekeredést felismerő rendszerek és a fehérje lebontó folyamatok. Valószínűsíthető az is, hogy ezek a fehérje szakaszok valamilyen fontos funkciót is hordoznak, ennek köszönhető, hogy az evolúciósan hátrányos amiloidogén tulajdonság nem szelektálódhatott ki.

Az ismert amiloid képző fehérjeszakaszok felhasználásával több algoritmus készült már a potenciálisan amiloidogén szakaszok felismerésére. A módszerek részben ezen algoritmusok közül a legtöbbet használt változatokat részletesen ismertetem majd, de lényegében minden program az elsődleges szerkezetből, vagyis az aminosav szekvenciából próbálja megjósolni a potenciálisan amiloid képzésre hajlamos fehérjeszakaszokat. Ezek a módszerek tehát vagy ismert amiloid képző szekvenciák alapján vagy in vitro tesztelt rövid peptid szekvenciák alapján lettek „betanítva” arra, hogy a béta-lemez struktúrába nagy valószínűséggel beépülő szekvenciákat megtalálják. Az ilyen módszereknél az egyik fő veszély a magas fals pozitívítási arány, hiszen a felismeréshez használt ismert amiloidogén szakaszok mellett viszonylag kevés információ áll rendelkezésünkre a bizonyítottan nem amiloid képző szekvenciákról. Természetesen előállítható sok olyan random rövid peptid szakasz, melyeken az amiloid képződés in vitro tanulmányozható (a későbbiekben ismertetésre kerülő WALTZ programnál ezt alkalmazták), ezzel azonban csak arra kapunk választ, hogy egymagában egy ilyen szakasz mennyire amiloidogén. A szervezeten, sejten belül természetesen az amiloidogén szakaszok egy fehérjeláncon belül találhatók, és akár annyira rejtve, szerkezetbe foglalva is lehetnek, hogy sosem lesz alkalmuk egy másik, esetünkben kóros amiloid szerkezetbe beépülniük. Az amiloid képződés megértésében és a pontosabb jósló programok kifejlesztésében nagy szükség lenne arra, hogy a kérdéses szakasz közvetlen környezetét is minél pontosabban feltérképezzük, valamint megértsük a motívumban és az azt határoló „flanking” régióban munkálkodó fizikai és kémiai tényezőket, melyek amiloid szerkezet kialakulásához vezetnek.

2. Célkitűzések

Az MTA TTK Enzimológiai Intézetében végzett munkám egy nagyobb átfogó bioinformatikai kutatás része, ezért célkitűzéseimet két részben ismertetem.

A dolgozatomban ismertetett munkám fő célja az amiloid képződés hátterének részletes megismerése, a kóros szerkezeti átalakulás szerkezeti alapjainak és bioinformatikai vizsgálati lehetőségeinek megismerése volt. Céлом volt továbbá az összes kísérletesen igazolt amiloidogén szekvencia összegyűjtése, a rendelkezésre álló algoritmusok általi potenciális amiloidogén szekvenciák jóslása és összegyűjtése. Az itt ismertetett munkámban az adatgyűjtésen kívül elvégeztem a szekvenciák elsődleges kiértékelését, melyek alapul szolgálnak a kutatás további elemzéseihhez.

Bár a dolgozatomnak nem része, de fontosnak tartom megemlíteni a kutatásunk hosszú távú céljait, hiszen elsősorban ezek adnak értelmet a munkámnak. A laborban folyó átfogó bioinformatikai vizsgálat fő kérdése az, hogy a rendezetlen fehérjeszakaszokon belül található fehérje-fehérje kapcsolatban résztvevő motívumok és azok közvetlen környezete milyen fizikai és kémiai tulajdonságokkal rendelkeznek, mik azok a kulcsfaktorok, amik ezeket a motívumokat egymástól elkülönítik. Az egész vizsgálatnak nagyon komoly evolúciós vetülete is van, arra is keressük a választ, hogy miként lehet jó interakciós motívum egy rövid rendezetlen fehérjeszakasz anélkül, hogy potenciális amiloid képző szakasszá váljon. A vizsgálat alapjait a dolgozatomban ismertetett módszerek képezik, első megközelítésben tehát a bioinformatika eszközeit használjuk, de a későbbiekben mindenképpen kísérletes úton is igyekszünk bizonyítani az *in silico* kapott eredményeinket.

3. Módszerek

3.1. Adatgyűjtés

Az amiloid képző szakaszok vizsgálatában két forrásra, kísérleti eredményekből származó illetve szekvencia alapján jóslt adatokra támaszkodtam.

A kísérleti adatokat a szakirodalomból és az ahhoz kapcsolódó adatbázisokból gyűjtöttem össze. Több in vitro kísérletben csak a konkrét amiloid kötő szakaszt vizsgálták, ezekből kizárólag az amiloidogén peptid szakaszokat tudtam volna kinyerni. Mivel elsődleges célom az amiloidogén régió és a környezetének részletes vizsgálata, ezért a szakirodalom kutatásom olyan irányba fordult, ahol konkrét ismert akár patológiás fehérjéket vizsgáltak meg, és ahol megállapították az amiloidogén szakaszt. Ezen adatok segítségével és a Uniprot adatbázis felhasználásával készült el az ismert amiloid képző régiókat tartalmazó adattáblám, mely a kötő és az előtte illetve utána lévő ún. flanking szekvenciákat tartalmazza. Mivel munkám elsősorban a patológiás amiloidózis hátterének felderítésére irányul, a kísérletes munkák közül kiszűrtem azokat az amiloidogén régiókat, melyek fiziológiás amiloid képzésben vesznek részt (ezek elsősorban élesztő fehérjék). A kísérleti eredményekből nyert amiloidogén szakaszok listáját az 1. táblázat tartalmazza.

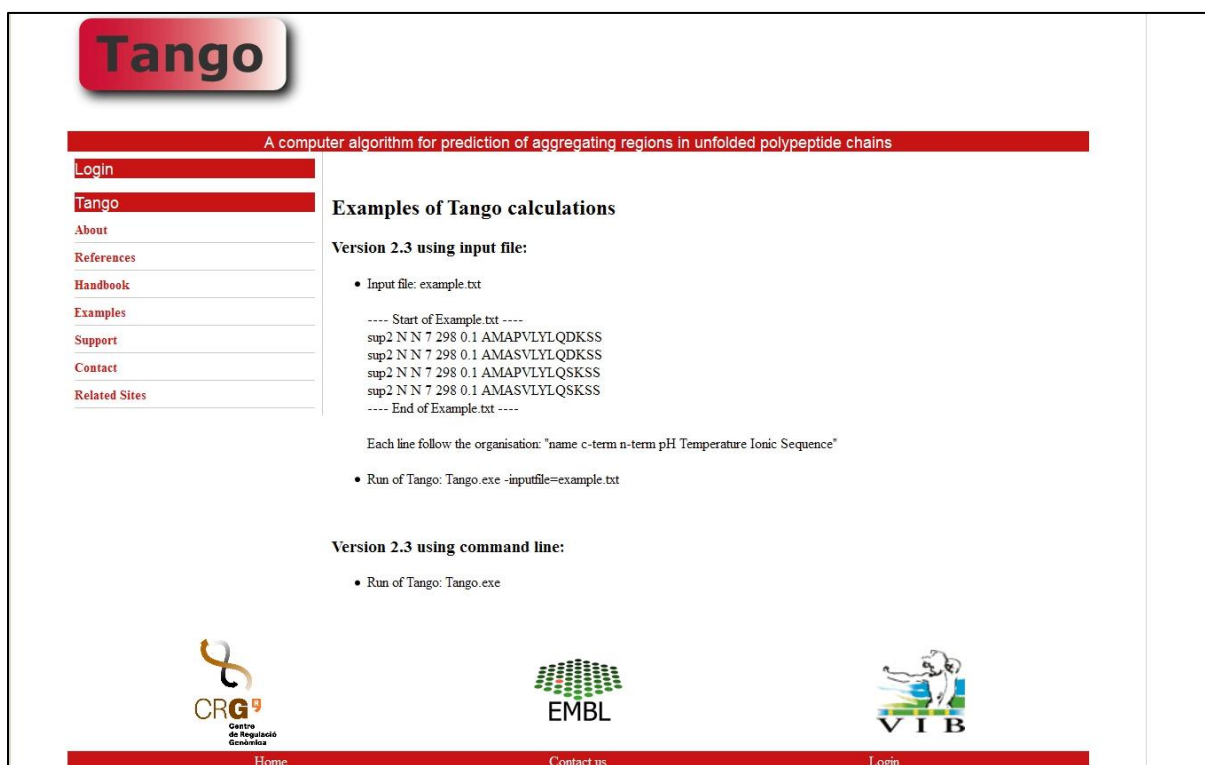
1. Táblázat. Az analízis során felhasznált kísérletesen igazolt amiloidogén régiók listája [14,15]

UniProtAC	Elnevezés:	Faj:	Szekvencia:
Q86WK6	AMIGO1	H. sapiens	LLPSLSLLFEVARA
P02655	Apolipoprotein C-II	H. sapiens	NLYEKTYLPAVDEKLRDL
P61769	Beta-2-microglobulin	H. sapiens	IQRTPKIQVYSRHPAENGKSN
P61769	Beta-2-microglobulin	H. sapiens	LLKNGERIEKVEHSDLSFSKD
P02671	Fibrinogen alpha chain	H. sapiens	VVTSEDGSDCPEAMDGLTSGIG
P02724	Glycophorin-A	H. sapiens	VYPPEETGERVQLAHHFSEPEITLIIFG
P10636	Tau protein	H. sapiens	FTFHVEITPNVQKEQAHSEE
Q08431	Lactadherin	H. sapiens	GDVFPSYT
P11686	SFTPC	H. sapiens	FGIPCCPVHLKRLIVVVVVVLIVVVIVGALLMGL
P02766	Transthyretin	H. sapiens	CLAGLVFVSEA
Q7Z6G8	ANKS1B	H. sapiens	ALVEK
Q7Z6G8	ANKS1B	H. sapiens	LGGGSG
Q7Z6G8	ANKS1B	H. sapiens	LPLSN
P02671	Fibrinogen alpha chain	H. sapiens	SGIGTL
P02766	Transthyretin	H. sapiens	SPFHEHAEVVF
P04156	Major prion protein	H. sapiens	SNQNNF
P10636	Tau protein	H. sapiens	VQIVYK
P01308	Insulin	H. sapiens	VEALYL
P01308	Insulin	H. sapiens	LYQLEN
P40967	Melanocyte protein PMEL	H. sapiens	GQVIWVN
P40967	Melanocyte protein PMEL	H. sapiens	AEVSIVV
P40967	Melanocyte protein PMEL	H. sapiens	LASLIYR
P07184	Chorion protein S18	D. melanogaster	DAQAIAL
P17110	Chorion protein S36	C. capitata	RQGNINIVA
P17110	Chorion protein S36	C. capitata	VQQNYQA
P07182	Chorion protein S36	D. melanogaster	QQEVINK
Q8JOV5	Hydrophobin	N. fumigata	GLINIAT
O60048	Hydrophobin	P. ostreatus	FNGLIVV
Q8WZI5	Hydrophobin 1	P. ostreatus	GILNIVV
Q8WZI6	Hydrophobin 1	P. ostreatus	SLLNIVV
Q8WZJ2	Hydrophobin 1	D. glabratum	TAFTILA
Q8WZJ3	Hydrophobin 1	D. glabratum	GVVAIGC
Q6YF32	Hydrophobin 1	D. glabratum	NLSLFDQ
Q6YF31	Hydrophobin 2	G. moniliformis	TVIAFLA
Q9P8T0	Hydrophobin 1	L. edodes	YGNLISL
Q6YF29	Hydrophobin 4	G. moniliformis	EFQEICA
P28307	Major curlin subunit	E. coli	AIAAIVF
P28307	Major curlin subunit	E. coli	SELNIYQY
P0A1E5	Major curlin subunit	S. typhimurium	TLSIYQY
B1W2X5	Chaplin	S. griseus	GNTCVN

A kísérletes adatok mellett a további analíziseinket in silico jósló amiloidogén szekvenciákon is elvégeztük. A predikcióhoz szükséges szekvenciákat a Uniprot adatbázisból töltöttem le, melyekből a redundancia elkerülése érdekében csak a humán fehérjéket szűrtem ki. A kiválogatás után összesen 20235 fehérjéből álló szekvencia adatbázis fájlt hoztam létre, amit a kutatócsoport szerverén tárolok, és melyen különböző jósló szoftvereket futattam.

Két amiloid jósló algoritmust használtam fel, a CRG-EMBL-VIB által fejlesztett TANGO aggregálódó és fel nem tekeredett láncok felismerésére kifejlesztett algoritmust használtam, saját szerveren futtatva. Több beállítási lehetőség közül, a kiválasztott program beállításokkal ($nt="N"$ $ct="N"$ $ph="7.4"$ $te="303"$ $io="0.05"$ $tf="0"$ $stab=-10$) futtatva 1013 fehérjében összesen 1743 amiloid képző szakaszt jóslt.

A 2004 októberében publikált TANGO aggregációs és fel nem tekeredett láncokat jósló algoritmus, mely a β -redő fizikai-kémiai tulajdonságain alapul, kiegészítve azzal a feltételezéssel, hogy az aggregátum központi régiói teljesen elásottak a fehérjében (5. ábra) [16].



5. ábra. A TANGO szoftver honlapja, példa kimeneti fájlal.(képernyő mentés)

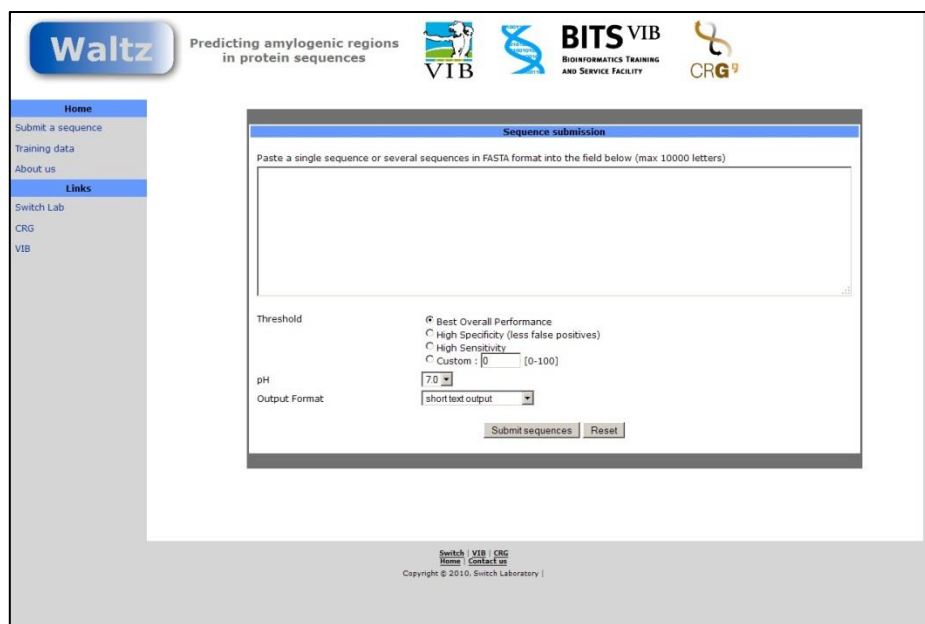
Az algoritmus alapja a szakirodalomból származó 179 és további 71 új (köztük például: prion protein, lizozim és β 2-mikroglobulin) peptid szekvencia. Mivel a kísérletek azt mutatták, hogy bizonyos fehérje szekvenciák elősegítik az aggregációt és az amiloid képződést, ezért beépítettek egy statisztikai mechanizmust, ami azonosítja a fehérjében a β -aggregáló

szakaszokat. A TANGO algoritmusnál különböző környezeti paraméterek is megadhatók, úgy mint a pH, koncentráció, ionerősség és trifluoretanol koncentráció. A TANGO négy különböző konformációs állapotot és különböző energia szinteket vizsgál meg, figyelembe véve a hidrofóbicitást, szolvatációs energetikát, elektrosztatikus kölcsönhatásokat és a hidrogén-híd kötést.

A TANGO működése:

- Az aminosav szekvencia α -hélix hajlamának vizsgálatára az AGADIR-1s paramétert használja a program, kétablakos megközelítést beépítve [17].
- β -turn hajlamot 3 energia adatból számolja: (1) β -turn kompatibilitási konformációs entrópiájának aminosav specifikus értéke (2) minden aminosavnak turn struktúrával való pozíciófüggő interakciója (3) az i és i+3 láncpozíció közti H-híd kötés kialakulásának valószínűsége.
- β -redő aggregációjához a következő feltevésekkel él: (1) rendezett β -redő aggregátum másodlagos szerkezete (2) Az aggregációban érintett régiók teljesen el vannak ásva/rejtve. (3) az ellentétes töltések egy specifikus ablakon belül csökkentik az aggregációs készséget.
- Az amiloid képződésre ható fizikai-kémiai hatások, mint a pH, hőmérséklet, ionerősség, elektrosztatikus kölcsönhatások figyelembe vétele.
- Trifluoretanol hatásának vizsgálata (40%-os koncentrációig lineáris).

Ugyancsak a VIB gondozásában készült WALTZ amiloidogén szakaszt jósló szoftvert is használtam [18]. Több beállítási lehetőség közül a pH-t: 7.0-ra a predikciós határértéket: 97%-osra állítottam. A vizsgált 20235 szekvenciából 831 fehérjében 2780 amiloid képző régiót talált a program (6. ábra).



6. ábra. A WALTZ amiloid jósló program webes felületének nyitóoldala (képernyő mentés)

A WALTZ szoftver készítésekor több mint 200 kísérletileg is bizonyított peptid szekvenciát vettek alapul. A program készítésének idején a szekvencia adatok közül elsősorban hexapeptidek amiloid képző hajlamából szűrtek le információt. Egyik ilyen adatbázis az AmylHex, ami összesen 67 db amiloid képzésre nézve "pozitív" és 91 "negatív" hexapeptidből áll. Az amiloidogén 'STVIIIE' szakasz mutánsainak érdekes módon igen jelentős (51%-os) a túltrepezentáltsága. Hogy adatbázisuk a természetben előforduló amiloidogén szakaszok tulajdonságait jobban megközelítse, készítettek további 46 amiloid képző és 71 nem-amiloid képző peptid szekvenciát. A peptideket és a képződő aggregátumokat különböző módokon vizsgálták: elektronmikroszkóp, FTIR, röntgen diffrakció, és CD módszerekkel. Az amiloidogén hexapeptidek tulajdonságainak vizsgálatára létrehoztak egy pozíció specifikus pontozási mátrixot (PSSM), ami a jósló algoritmus alapja.

A jósló algoritmus három komponensből számítja ki a szekvenciához tartozó értéket:

- A szekvenciához tartozó pozíció specifikus mátrixból kiszámolt érték a S_{profile}
- Az algoritmus által figyelembevett 19 fizikai paraméterből számolja ki a S_{physprop} értéket
- A harmadik összetevőt a strukturális modellből származó pozíció specifikus pseudoenergia mátrixból számolja ki (S_{struct}).

Mind 3 összetevőhöz tartozik egy súly, így egy szekvenciához tartozó összérték:

$$S_{\text{total}} = \alpha_{\text{profile}} * S_{\text{profile}} + \alpha_{\text{physprop}} * S_{\text{physprop}} + \alpha_{\text{struct}} * S_{\text{struct}}$$

Miután összehasonlították a WALTZ és a TANGO algoritmust, kiderült, hogy a TANGO alkalmasabb bármilyen típusú, de különösen a béta-aggregáció kimutatására, míg a WALTZ az aggregációs hajlamot kevésbé észleli, azonban sokkal jobban jósolja az amiloidogén aggregációs hajlamot.

3.2. Amiloidogén szakaszok feldolgozása

3.2.1. Aminosav kompozíció elemzése

Az adatgyűjtés során kapott amiloidogén és az azt határoló szakaszok kiértékelését az aminosav összetételük elemzésével kezdtem. A továbbiakban ahol lehetőségem volt, mindig a motívum előtti 20 aminosavat, a motívumot, és az utána következő 20 aminosavat elemeztem.

3.2.2. Aminosav index szerinti kategorizálás

A kísérleti és jósolt szekvenciák elemzésénél több aminosav hasonló tulajdonságai miatt nem csak arra voltam kíváncsi, hogy milyen pozícióban milyen aminosav jellemző, hanem inkább az adott pozícióban az aminosavak jellemző fizikai és kémiai, szerkezeti és egyéb tulajdonságait próbáltam felderíteni.

A csoportosításhoz az Institute for Chemical Research, Kyoto University kutatói által publikált AAindex adatbázist használtam, mely az aminosavaknak különböző fizikai- kémiai tulajdonságait tárolja [19,20]. Az AAindex adatbázis jelenleg 3 részből áll: az AAindex1-ben minden aminosavhoz az adott tulajdonság alapján egy számot rendeltek, az AAindex2-ben aminosav subsztitúciós mátrixok találhatók, míg az AAindex3 statisztikai fehérje kapcsolati potenciálokat tartalmaz. Az adatbázis letölthető az adatbázis FTP szerveréről (<ftp://ftp.genome.jp/pub/db/community/aaindex/>)

Munkám során az AAindex1-et használtam, mely összesen 544 különböző tulajdonság szerint ad értékeket az egyes aminosavaknak (aktuális 9.0-ás verzió). 1988-ban Nakai és munkatársai 222 indexet állítottak össze, melyet 4 fő osztályba soroltak: (1) α -helix és turn hajlam, (2) β -redő hajlam, (3) hidrofóbicitás és (4) egyéb. 1996 óta Tomii és Kanehisa tovább bővítette az adatbázist, így érte el mai formáját az AAindex1. Minden bejegyzéshez tartozik egy azonosító, egy rövid leírás, referencia információ és az aminosavak numerikus értéke. A bejegyzések PubMed és LitDB keresztreferenciákat is tartalmaznak (7. ábra).

```
//
H CIDH920102
D Normalized hydrophobicity scales for beta-proteins (Cid et al., 1992)
R LIT:1817105b PMID:1518784
A Cid, H., Bunster, M., Canales, M. and Gazitua, F.
T Hydrophobicity and structural classes in proteins
J Protein Engineering 5, 373-375 (1992)
C CIDH920105 0.969 ZHOH040101 0.939 CIDH920101 0.921
  BASU050102 0.914 CIDH920103 0.911 ZHOH040103 0.909
  CIDH920104 0.904 NISK860101 0.897 ROBB790101 0.896
  NOZY710101 0.889 MEEJ810101 0.887 PLIV810101 0.877
  LEVM760106 0.873 MIYS850101 0.873 CORJ870102 0.872
  WERD780101 0.871 SWER830101 0.870 ROSG850101 0.866
  BIOV880101 0.864 ZHOH040102 0.862 RADA880102 0.862
  ARGF820101 0.862 JOND750101 0.861 ROSM880104 0.859
  TAKK010101 0.859 BASU050101 0.858 MEEJ800102 0.856
  FAUJ830101 0.856 BASU050103 0.856 MEEJ810102 0.843
  BIOV880102 0.837 WIMW960101 0.837 RADA880108 0.833
  SIMZ760101 0.832 GUOD860101 0.831 GOLD730101 0.829
  PONF930101 0.820 EISD860101 0.819 ZASB820101 0.809
  LIFS790101 0.808 BLAS910101 0.805 CASG920101 0.802
  RACS770101 -0.825 VINM940103 -0.826 FUKS010103 -0.826
  GRAR740102 -0.842 MIYS990103 -0.845 BULH740101 -0.851
  GUYH850102 -0.855 MIYS990105 -0.859 MEIH800101 -0.867
  WOLS870101 -0.869 MIYS990102 -0.870 PARS000101 -0.871
  VINM940101 -0.872 MIYS990101 -0.872 KARP850101 -0.873
  MIYS990104 -0.877 OOBM770103 -0.877 GUYH850103 -0.904
  VINM940102 -0.925 PARJ860101 -0.930
I  A/L  R/K  N/M  D/F  C/P  Q/S  E/T  G/W  H/Y  I/V
  -0.08 -0.09 -0.70 -0.71 0.76 -0.40 -1.31 -0.84 0.43 1.39
  1.24 -0.09 1.27 1.53 -0.01 -0.93 -0.59 2.25 1.53 1.09
```

7. ábra. Példa bejegyzés az AAindex 1 adatbázisból. Az elemzéshez szükséges aminosavakra vonatkozó információt az alsó két sorból nyertem ki.

Hogy egyszerűsítsem az AAindex1-gyel való munkát, a letöltött adatbázisból készítettem egy olyan fájlt, ami minden sorban egy-egy AAindex azonosítóval kezdve, A R N D C Q E G H I L K M F P S T W Y V sorrendben tartalmazza szóközzel elválasztva az egyes aminosavakhoz tartozó index értéket.

Ezt az elkészített elemző programjaim felhasználják az in silico jósolt és a kísérleti úton bizonyított szekvenciákra is. Az adatok elemzéséhez használt legfőbb programom minden AAindex-hez megvizsgálja, hogy az adott pozícióban mi az átlagos érték. Mivel az AAindex legtöbb tulajdonsága más és más értékek között pontoz (van több, ami 0 és 1 közötti

értékekkel, de olyan is van, ahol az értékek 62,9 és 224,6 között vannak), ezért hogy az aaindexek egymással összehasonlíthatók legyenek, egy külön programban normalizáltam az indexeket. Az adott aaindex mimuma -5, a maximuma +5 értéket vesz fel, a többi értéket pedig arányosan. A kísérleti szekvenciák alacsony száma miatt a jobb statisztikai erő elérésének érdekében szekvenciális összevonást alkalmaztunk. Az amiloidogén szakasz előtti és utáni flanking régiót 3 aminosavanként összevontuk (így összesen 6-6 pozíció lett a két határoló régióból), az amiloidogén motívumot pedig 3 szakaszra osztottam, az első és utolsó három aminosav képzett 1-1 pozíciót illetve a fennmaradó közepe egy harmadikat. Az ebből számolt pozícióátlag számolást az alábbi képlettel is összefoglalhatjuk, ahol az s jelenti az adott aminosav adott aaindexéhez tartozó értéket, l a kiválasztott fehérje szekvenciát, p adott pozíciót a szekvenciában, n pedig az aminosavak száma:

$$x = \frac{\sum_l (\sum_q S_{lp})}{n}$$

Az így kapott átlagértékeket az előző programmal -5 és 5 közötti értékkel láttam el. Az így megkapott adatsorból válogattuk ki azokat az aaindex tulajdonságokat, amik elég informatívak a motívum és a határoló régiók elkülönítésében.

A kiválogatás során csoportosítottam is az egyes indexeket alapvető tulajdonságaik szerint: fizikai, fizikai-kémiai, flexibilitás, hidrofóbicitás, statisztikai, szerkezeti, szerkezeti β -redő, szerkezeti coil, szerkezeti hélix és szerkezeti turn alcsoportokat képeztem.

3.2.3. Aminosav index értékek statisztikai szűrése

A további munkánkhoz olyan aminosav indexekre lesz szükségünk, amelyek a legjobban, de ugyanakkor különböző tulajdonságok, megközelítések alapján különítik el az amiloidogén motívumot, a motívum flanking régióját egy véletlenszerűen választott szekvenciától. A korábban leírt, elsősorban vizuális értékelésen alapuló előszűrő módszer után az aminosav indexek alapján kapott pozíció értékeket statisztikai módszerekkel újraelemezzük, és a kapott

normál eloszlásbeli különbségek alapján szűkítjük tovább a felhasználni kívánt aminosav indexek körét. A módszerben a következő lépéseket alkalmazzuk:

- az aminosavak aaindex értékeit átlagolom a fragmentumon belül (fragmentum pl. a motívumot megelőző 3 aminosav).
- Az adatok ezután átesnek egy Z-transzformáción
- Az adatokat egy -1 - 1 range transzformációval uniformizálom
- Végül a normál eloszlás illesztése történik meg, Naïve Bayes klasszifikációval és Laplace-korrektcióval.

4. Eredmények

4.1. Aminosav kompozíció

Összehasonlítva a TANGO, WALTZ és a kísérletes összeállításból származó aminosav kompozíciókat, látható, hogy a jósvolt szekvenciáknál főleg csak a kötő motívumban van változás, az előtte és az utána lévő szakasz sokkal egységesebb képet ad. A kísérletes eredmények alapján számolt aminosav összetétel sokkal változatosabb képet ad (8. Ábra).

A	3	11,8	9,7	4,7	10,7	7,7	7,7	6,1	8,5	7	14,4	7,5	9	12,1	4
R	1	2,9	1,9	0,9	4,5	4,3	1,7	2	2,6	2,6	5,4	3,7	2	4	0
N	11,1	7,8	5,8	3,7	6,2	5,1	5,1	8,1	6,8	5,2	7,2	3,7	2	3	5,1
D	2	2	2,9	6,5	2,7	4,3	1,7	2,5	2,6	5,2	2,7	2,8	3	3	6,1
C	5,1	4,9	5,8	3,7	5,4	0,9	0,9	2	2,6	7	1,8	2,8	7	1	2
Q	6,1	2,9	3,9	4,7	2,7	0,9	10,3	2,5	3,4	0	4,5	1,9	5	1	4
E	3	4,9	3,9	9,3	1,8	4,3	4,3	9,6	2,6	1,7	0,9	1,9	4	5,1	8,1
G	13,1	15,7	12,6	10,3	11,6	13,7	11,1	7,1	6	12,2	9	11,2	7	10,1	10,1
H	0	2	3,9	4,7	0,9	0,9	0	3	0,9	1,7	1,8	2,8	3	1	2
I	3	2	4,9	5,6	4,5	6	6,8	5,6	12,8	3,5	4,5	8,4	5	5,1	4
L	5,1	6,9	4,9	7,5	10,7	12	17,1	15,2	9,4	9,6	12,6	10,3	6	7,1	8,1
K	6,1	4,9	3,9	6,5	7,1	3,4	0,9	3,5	4,3	2,6	5,4	1,9	3	4	4
M	2	1	1	0	2,7	0,9	0	0,5	0,9	0,9	0,9	1,9	1	2	0
F	2	0	1,9	0,9	0	1,7	4,3	2,5	5,1	2,6	1,8	4,7	2	5,1	1
P	3	3,9	6,8	5,6	6,2	7,7	2,6	5,1	0	8,7	2,7	5,6	10	8,1	8,1
S	7,1	11,8	12,6	10,3	10,7	7,7	6,8	6,6	6,8	14,8	8,1	11,2	6	9,1	9,1
T	8,1	5,9	2,9	1,9	6,2	5,1	4,3	3,5	2,6	6,1	6,3	3,7	13	7,1	13,1
W	1	0	0	0,9	0,9	0,9	0	0	0,9	0,9	0,9	0	2	1	1
Y	4	2,9	1,9	5,6	0	3,4	3,4	2,5	7,7	2,6	1,8	5,6	4	2	2
V	14,1	5,9	8,7	6,5	4,5	9,4	11,1	12,1	13,7	5,2	7,2	8,4	6	9,1	8,1

A	6,9	6,5	7,3	7,8	8,6	7,6	8,3	9,9	7,5	8,2	7	6,7	5,7	6	6,6	A	6,3	7,4	7,1	7,3	7,0	9,9	7,0	6,3	6,1	6,5	6,6	6,8	6,3	6,3	7,0
R	5,7	5,6	6,3	5,5	3,9	1,2	0,1	0,1	0,1	3,7	6,1	7	6,6	6,4	5,7	R	5,1	4,9	4,5	4,8	4,7	4,7	0,2	1,0	2,9	5,6	5,2	4,7	5,2	5,6	5,0
N	3,1	4	3,9	3,4	2,9	1,4	0,5	1,2	0,4	3,3	3	3	3,8	4,2	4,2	N	3,9	4,0	3,6	3,8	3,9	4,7	5,9	5,6	5,2	4,1	4,4	3,4	3,9	4,2	3,9
D	3,9	3,7	3,6	3,2	2,2	0,9	0	0	0,1	1,8	2,8	3,2	3,9	4,3	4	D	4,6	4,3	4,6	4,5	4,7	3,6	2,2	2,3	3,8	5,5	4,7	4,8	4,3	4,6	4,8
C	2	2,4	1,9	2	2,7	1,6	2,3	3,3	1,8	3	2,6	2,8	2,3	2,3	2,1	C	2,2	2,1	2,1	2,4	2,3	1,6	1,3	1,7	2,1	2,9	2,1	2,0	1,9	2,6	2,0
Q	3,7	4	3,9	3,6	2,1	0,8	0,4	0,5	0,4	3,1	3,9	4,5	4,3	4,3	4	Q	4,2	3,7	4,0	3,8	3,6	3,7	5,7	3,8	3,3	4,3	3,8	4,2	4,2	3,9	4,3
E	5,7	4,4	3,9	3,7	2,7	1	0,1	0,3	0,2	3,1	3,3	4,6	4,8	5,6	6,6	E	6,2	5,7	5,3	5,7	5,5	4,7	2,8	2,6	4,1	6,2	6,0	5,7	5,9	5,8	5,9
G	6,2	6,2	6,5	6,8	7,3	4,4	3,7	4,7	2,2	7,4	7,2	6,4	4,8	5,8	6	G	6,5	6,0	5,6	6,1	6,4	8,6	1,2	2,7	5,3	7,6	6,5	6,3	5,1	6,2	5,6
H	2,5	2,1	2,1	2,1	1,7	0,6	0,2	0,4	0,4	2	2,8	2,3	3,1	2,5	2,4	H	2,4	2,5	2,2	2,6	2,6	2,8	0,5	0,8	1,5	2,8	2,7	2,5	2,9	2,6	2,9
I	5,5	4,8	5,7	6,6	7,7	12,2	12,8	12,8	14,9	6,9	6	5,5	5	5,5	5,9	I	5,7	5,4	5,9	5,5	5,3	5,1	10,0	15,0	11,5	3,7	5,2	6,1	5,6	5,3	5,6
L	11,1	10,7	11,5	12,3	13,5	21,9	28,2	22,6	23,7	14,8	11,7	9,7	10,9	10,1	11,1	L	11,2	11,4	11,7	11,0	11,9	11,3	20,0	12,1	8,7	10,9	10,8	11,6	10,7	11,1	11,3
K	4,5	5,4	5,6	5,2	3,4	1	0	0,3	0,2	2,8	5,1	6,6	6,3	5,5	5,6	K	5,2	5,6	5,6	5,6	5,5	5,2	0,3	1,2	2,9	5,9	6,4	5,7	5,6	5,6	5,8
M	3	3,4	2,5	3,4	3,9	2,8	2,6	3	2,3	3,1	2,5	2,4	1,9	2,2	2,3	M	2,5	2,7	2,8	2,5	2,3	2,5	0,6	1,8	2,7	2,5	2,6	2,3	2,6	2,4	2,2
F	4,5	5	5,1	5,3	5,8	11,3	12,1	11,4	14,3	5,8	5,8	5,5	5	5	4,7	F	5,0	4,7	4,9	5,0	4,3	3,9	12,0	12,1	8,1	3,1	4,2	4,9	5,1	4,9	4,3
P	5,9	6,2	5	4,6	5,5	1,7	0	0	0	3	4,9	4,9	5,5	6,1	5,7	P	4,6	4,7	5,6	5,2	4,7	4,2	0,0	1,2	3,7	5,3	5,4	4,6	5,2	4,6	5,3
S	7,5	8,3	7,8	7,3	6,9	4,3	3,3	4,6	2,3	7,8	6,9	7,4	7	7,4	7,1	S	6,7	6,5	6,8	7,0	7,7	6,3	3,1	4,7	6,6	7,7	6,8	6,7	7,6	6,6	7,2
T	6	5,4	5,4	5,1	5,9	4,3	5,1	5,4	3	6,2	4,6	5,3	5,5	5,8	5,3	T	5,6	6,2	5,5	5,9	5,2	5,9	3,0	3,4	4,9	5,3	5,1	5,5	5,6	5,2	5,5
W	1,8	2,4	2,6	2,2	2	2,6	2,2	1,8	3,1	2,3	2,6	2,1	2,1	2,1	2	W	1,6	1,8	1,6	1,4	1,5	1,5	4,3	2,5	1,0	1,1	1,5	1,5	1,6	1,8	1,3
Y	3,2	3,7	3	3,2	3,4	3,3	2,4	2,9	4,6	3,6	3,8	4	4,3	3,5	3,2	Y	3,7	3,5	3,2	2,9	3,4	2,9	6,5	10,5	9,1	2,1	3,0	3,6	4,3	3,8	3,3
V	7,2	5,9	6,2	6,8	7,8	15	15,6	14,9	18,7	8,2	7,4	6,1	7,2	5,2	5,8	V	6,8	7,1	7,3	7,0	7,6	6,8	13,2	8,7	6,4	6,8	6,8	7,1	6,3	6,9	6,6

8. ábra. Az amiloidogén szekvenciák és környezetük aminosav kompozíciója. A felső ábrán a kísérletesen bizonyított, míg alul a TANGO (jobbra) és WALTZ (balra) programokkal jósvolt amiloidogén motívumok (fekete keretben) és környezetük aminosav kompozícióját láthatjuk. Az egyes cellákban található gyakoriságok nem egyes aminosavakra vonatkoznak, hanem összevont átlagok (lásd módszerek fejezet)

A hidrofóbb aminosavak (F, I, L, V, A) motívumon belüli dúsulását egyértelműen meg lehet figyelni. Érdekes megfigyelni, hogy míg a jósolt szekvenciákban az amiloidogén motívumban a hidrofób aminosavak aránya egyformán magas, a kísérletesen bizonyított szekvenciákban néhol aszimmetria észlelhető (pl. leucin és izoleucin esetében).

A prolin esetében szintén egyértelmű a tapasztalt alacsony arány a motívumon belül, hiszen ez az aminosav igen nehezen épül be bármilyen másodlagos szerkezeti elembe, márpedig az amiloid fibrillum alapvető szerkezeti eleme a másodlagos béta-redő szerkezeti elem. Ugyanakkor figyelemreméltó a prolin mennyiségének növekedése a határoló flanking régiókban, ezeken a helyeken szerepe lehet az utána következő, vagy az előtte lévő másodlagos szerkezeti elem stabilizálásában.

A töltött aminosavak szintén kis gyakorisággal fordulnak elő az amiloidogén motívumon belül, ezek az aminosavak elsősorban a rendezetlen vagy az alfa-hélix struktúráknak kedveznek, és magas hidrofilitásuk miatt kifejezetten csökkentik az aggregációs hajlamot. Fontos megjegyezni, hogy a motívumtól távolabb eső határoló régiókban a töltött aminosavak aránya igen magas (hiszen ezek sokszor rendezetlen szakaszok), de a motívum felé közeledve már a flanking régióban is igen alacsony az előfordulási valószínűségük (különösen igaz ez a lizin és a glutaminsav esetében).

A szerin, mint erősen hidrofil aminosav aránya szintén alacsony az amiloidogén motívumon belül, de a korábban említett töltött aminosavakkal ellentétben a környező flanking régiókban mindenhol magas az előfordulási gyakorisága.

A glicin a legkisebb és legmozgékonyabb aminosav, így szintén jól magyarázható magas aránya az általában flexibilis rendezetlen határoló régióban, és aránylag alacsony gyakorisága merev motívumon belül.

A rendezetlen fehérjékben található kötőmotívumok esetében a foszforilálható aminosavak kiemelt szerepet kapnak, hiszen ezek poszttranszlációs módosításával a kötődés erőssége jelentősen változhat. Az amiloidogén motívum környezetében található ezen aminosavaknak szintén lehet szerepe a foszforilációban, az így bevitt töltés pedig erősen befolyásolhatja a motívum amiloid képző hajlamát. További érdekesség a kísérletes eredményekből származó

A glutamin előfordulásában is látható egy érdekes jelenség, erőteljes feldúsulást láthatunk mind a kísérletes, mind a WALTZ által jósolt amiloidogén motívumok N-terminális végén.

Mivel több AAindex kiértékelés nem mutatott érdemi információt (nem volt látható semmilyen elkülönülés az amiloidogén motívum és a határolórégiók között), ezért az 544 tulajdonság közül első körben összesen 219 tulajdonságot választottam ki. A kiválasztáshoz a régiókat a módszerek fejezetben leírtak alapján felbontottam, a rájuk vonatkozó aminosav index átlagokat uniformizáltam, és excel program segítségével, feltételes formázás (színezés) után vizuálisan válogattam ki az informatívnak vélt indexeket (9. ábra). A kiválasztott AAindexeket 11 kategóriába soroltam, melyeken belül majd a következő fejezetben bemutatandó statisztikai kiértékelés segítségével fogom kiválasztani a leginformatívabbakat. A kategóriákon belüli további szűkítést majd akkor tudom biztonságosan megtenni, ha már a többi típusú kötőmotívumokról is rendelkezésünkre állnak majd az AAindex információk.

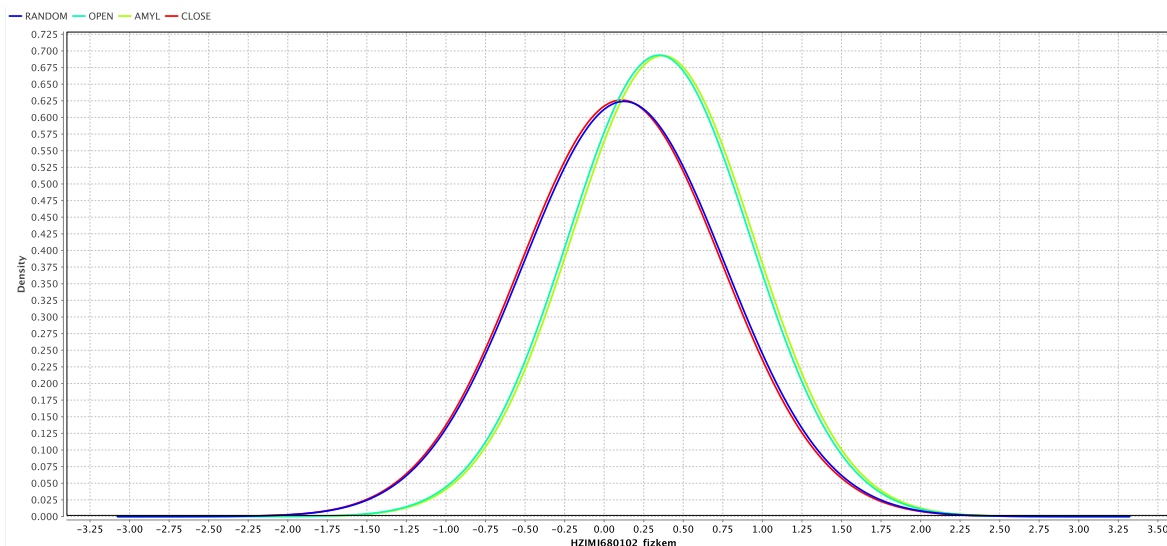
HBIGC670101	-1	-2	-1	-1	-1	-1	0	0	0	-1	-1	-1	-1	-1	HBIGC670101	0	0	0	0	0	1	0	1	0	0	0	0	0	HBIGC670101	0	0	0	0	0	1	1	0	-1	0	0	0	0	
HFAUJ880102	1	1	1	1	1	0	2	2	3	1	2	1	1	1	HFAUJ880102	2	2	2	2	3	3	4	2	2	2	2	2	2	HFAUJ880102	2	2	2	2	2	3	3	2	2	2	2	2	2	
HLEVM760106	-1	-1	-1	0	1	0	0	0	-1	1	-1	0	0	-1	HLEVM760106	0	0	0	0	2	2	1	2	0	0	0	0	0	HLEVM760106	2	1	1	0	0	0	0	0	2	1	1	0	0	0
HFASG760102	0	0	0	0	0	1	1	1	1	0	1	1	0	0	HFASG760102	0	0	0	1	2	2	2	2	1	1	0	0	0	HFASG760102	1	1	1	0	1	1	2	1	0	0	1	0	0	
HFAUJ880111	5	4	4	4	4	5	-5	-5	-5	-5	4	4	5	-5	HFAUJ880111	4	4	4	4	5	-5	-5	-5	4	4	4	4	-4	HFAUJ880111	4	4	4	4	4	-5	-5	-5	4	4	4	4	4	
HZIMJ680102	1	0	1	1	1	1	2	2	2	1	1	1	1	1	HZIMJ680102	2	2	2	2	3	4	3	4	2	2	2	2	2	HZIMJ680102	2	2	2	2	1	3	3	2	1	2	2	2	2	
HNOZV710101	-4	-4	-4	-4	-3	-3	-3	-2	0	-3	-3	-3	-3	-4	HNOZV710101	-3	-3	-3	-3	-3	-1	1	0	-2	-3	-3	-3	-3	HNOZV710101	-3	-3	-3	-3	-3	-1	-1	-2	-3	-3	-3	-3	-3	
HMIY5990102	1	2	2	2	2	1	0	0	1	1	0	1	1	2	HMIY5990102	1	1	1	0	-2	-1	-2	0	0	0	1	1	1	HMIY5990102	1	1	1	1	1	-2	-1	0	1	1	1	1	1	
HMIY5990103	0	0	0	0	0	0	-1	0	0	0	-1	0	0	1	HMIY5990103	0	0	0	-1	-3	-4	-3	-4	-1	0	0	0	0	HMIY5990103	0	0	0	0	0	-3	-2	-1	0	0	0	0	0	
HKARP850101	1	1	1	1	1	0	0	-1	0	0	-1	0	0	0	HKARP850101	0	0	0	0	-2	-2	-3	-1	0	0	0	0	0	HKARP850101	0	0	0	0	0	-2	-2	-1	-1	0	-1	0	0	0
HVINM940102	2	2	2	2	2	2	1	1	0	2	2	1	1	2	HVINM940102	1	1	1	1	-1	-1	-2	1	1	1	1	1	1	HVINM940102	1	1	1	1	1	-1	-1	0	1	1	1	1	1	
HVINM940103	1	1	1	1	1	1	0	1	1	0	1	0	1	1	HVINM940103	1	1	1	0	-1	-2	-1	-2	0	1	1	1	1	HVINM940103	1	1	1	1	1	-1	-1	0	0	0	1	1	1	
HCIDH920101	0	-1	0	0	0	0	1	1	0	0	1	0	0	-1	HCIDH920101	0	0	0	1	3	3	3	3	1	1	0	1	0	HCIDH920101	0	0	0	0	0	2	2	1	0	0	0	0	0	
HLAWF840101	-2	-2	-2	-2	-1	-1	-1	-1	-2	-1	-1	-1	-1	-2	HLAWF840101	-1	-1	-1	-1	0	0	0	0	-1	-1	-1	-1	-1	HLAWF840101	-1	-1	-1	-1	-2	0	-1	-2	-1	-1	-1	-1	-1	
HWILM950102	-2	-2	-2	-2	-2	-1	-1	-1	-2	-1	-1	-2	-1	-2	HWILM950102	-1	-1	-1	-1	0	1	1	1	-1	-1	-1	-2	-1	HWILM950102	-1	-1	-1	-1	-1	0	0	-1	-2	-2	-1	-1	-1	
HGRAR740101	-3	-3	-3	-3	-3	-4	-4	-2	-3	-3	-3	-3	-3	-3	HGRAR740101	-3	-3	-4	-4	-5	-5	-4	-5	-4	-3	-3	-3	-3	HGRAR740101	-3	-3	-3	-3	-3	-4	-4	-4	-3	-3	-3	-3	-3	
HVENT840101	-2	-2	-3	-3	-2	-1	0	0	-1	0	-1	-2	-3	-2	HVENT840101	-2	-2	-2	-2	2	3	3	3	-1	-1	-2	-2	-2	HVENT840101	-2	-2	-2	-2	-2	2	1	-1	-2	-2	-2	-2	-2	
HZHOH040102	-1	-2	-1	-1	-1	-1	0	0	0	-1	-1	0	-1	-1	HZHOH040102	0	0	0	0	1	2	0	1	2	0	0	0	0	HZHOH040102	0	0	0	0	-1	1	1	0	-1	0	0	0	0	
HCHOP780101	0	1	1	0	0	0	-1	-1	-1	0	0	0	0	0	HCHOP780101	-1	0	1	-1	-3	-3	-3	-3	-1	-1	-1	0	-1	HCHOP780101	-1	0	-1	-1	-1	-2	-2	0	-1	-1	-1	-1	-1	
HQIAN880107	1	1	1	1	1	0	2	2	2	0	2	1	1	0	HQIAN880107	1	1	1	2	2	3	3	3	2	2	1	1	1	HQIAN880107	1	2	2	1	2	2	2	1	1	2	1	1	1	1
HQIAN880116	1	1	1	1	1	0	0	0	0	1	0	1	1	1	HQIAN880116	0	0	0	0	-1	-2	-1	0	0	0	0	0	0	HQIAN880116	0	0	0	0	0	-1	-1	0	0	0	0	0	0	
HMIY5990105	0	0	0	0	0	-1	-1	-1	-2	0	0	-1	0	0	HMIY5990105	-1	-1	-1	-1	-3	-4	-3	-4	-1	-1	0	0	0	HMIY5990105	-1	-1	-1	0	-1	-3	-2	-2	-1	0	-1	0	-1	

A kiszűrt 219 tulajdonságot a következő módon soroltam kategóriákba. A fizikai tulajdonságok csoportjába összesen 6 tulajdonság került, mint például molekula méret. A fizikokémiai tulajdonságok csoportjába összesen 29 indexet soroltam, melyek közt szerepel információ az olvadásponttal, töltéssel, tömeggel kapcsolatban. Mindhárom a flexibilitás kategóriába sorolt indexnél megmutatkozik, hogy a kötő motívum merevebb szerkezetű, mint az azt határoló flanking régió. Külön kategóriába kerültek a hidrofóbicitásra vonatkozó indexek (34 db AAindex), ezekben különböző mértékben ugyan, de markánsan előtűnik az amiloidogén hidrofób karaktere, főként az előtte és mögötte elhelyezkedő határoló régiókhoz képest. Összesen 37 darab elemet tartalmaz a statisztikai alapokat használó indexek kategóriája, melyben az igen sokrétű index típus miatt sokféle mintázat figyelhető meg. A szerkezeti elemek kialakítására való hajlammal kapcsolatban az egyes másodlagos szerkezeti elemeknek megfelelően alkategóriákat képeztem. Ennek megfelelően 27 index a béta-redő szerkezettel, 25 index az alfa-hélix szerkezettel, 19 index a visszafordító „turn” szerkezettel és 9 darab a rendezetlen szerkezettel kapcsolatos. 23 szerkezetre vonatkozó indexet nem tudtam a fenti alkategóriákba sorolni, ezeket külön, általános szerkezeti kategóriaként kezelem. A leírás lapján egyik kategóriába sem sorolható indexet egy egyéb kategóriába soroltam (összesen 7db). Az összes 219 kiszűrt aminosav index azonosítóját és rövid leírását kategóriákba foglalva a melléklet tartalmazza.

4.3. AAindex eloszlások statisztikai feldolgozása

Az aminosav indexekből kapott nagy mennyiségű gyakoriság adat automatikus feldolgozására, kiértékelésére, és azok információtartalmának meghatározására statisztikai feldolgozó módszereket alkalmazok. Ezzel a módszerrel a jövőben a vizuális szűréssel alkalmasnak talált AAindexek körét fogom tovább szűkíteni több szempontot is figyelembe véve. Elsődleges szempont az eredményként kapott normál eloszlások különbözősége. A különbséget azonban nem kizárólag az amiloidogén motívumra nézve keressük, legalább annyira fontosak lesznek számunkra azok az AAindex eredmények, melyek akár csak az egyik közeli határoló régióra informatívak, vagy éppen a motívumra és vele együtt valamelyik közeli határoló régiójára. Bár láthatóan egy-egy aminosav indexből származtatott eredmény, a

különböző régiókra jellemző normál eloszlás markánsan nem különül el, de a végső analízisnél (ahol már ezek segítségével a többi rendezetlen kötőmotívumhoz hasonlítjuk ezeket) a több kis mértékben informatív index együttes használatával megfelelő eredményeket érhetünk el. A statisztikai feldolgozást követő normál eloszlás különbségekre mutat egy példát a 10. ábra.



10. ábra. Normál eloszlásbeli különbségek az amiloid képző motívumban és környezetében az aminosavak méretére vonatkozó aminosav indexből származó eredmények alapján. Az ábrán látható példában az AAindex értékek eloszlása a motívum után közvetlenül következő határoló régió (piros szín) a teljesen random szekvenciára hasonlít (kék vonal), az amiloidogén szakaszt N-terminális oldalról határoló régió (ciánkék vonal) pedig inkább a motívumnak (zöld vonal) megfelelő értékeket vesz fel.

5. Diskusszió

Az amiloidózishoz köthető betegségek patomechanizmusának felismerése óta illetve a háttérben álló kóros fehérje lerakódások közös szerkezeti tulajdonságának ismerete óta a tudomány számára egy lényeges kérdés, hogy a jelenség milyen mértékben terjeszthető ki a már ismert, de még amiloidózishoz nem köthető fehérjék körére. A kérdést in vitro megközelíteni szinte lehetetlen, hiszen a humán proteom több tízezer fehérjéjének szisztematikus megvizsgálása amiloid képző hajlam szempontjából jelenleg kivitelezhetetlen feladat. A potenciális új amiloidogén fehérjék megtalálása, a lehetséges célpontok számának feldolgozható méretűvé csökkentése a bioinformatikai kutatások, alkalmazások feladata. Erre a célra számos szoftver készült már több tudományos műhelyben is, melyek közül a két legtöbbet használt eszközt (TANGO és WALTZ) én is bemutattam dolgozatomban. Ezek a programok kiválóan alkalmasak arra, hogy egy ismert amiloidot képező fehérjeláncban lokalizálják, megtalálják a kóros szerkezet kialakításáért felelős szakaszt, esetleg szakaszokat, azonban igen nagy fals pozitív rátával működnek ismeretlen szekvenciák esetében. A dolgozatom egyik célja és a kapott eredmények alapján az egyik haszna az, hogy az eddig kevésbé vizsgált amiloidogén motívumot határoló régiókat és különösen a határoló régió – motívum átmenetet részletesebben ismerhetjük meg, feltárhatjuk azokat a tényezőket, melyek elősegítik, lehetővé teszik vagy éppen gátolják egy bizonyos régió belüli amiloidogén motívum kialakulását. A kapott eredményeken már most markánsan előtűnnek olyan aminosav tulajdonságok, melyek nem kizárólag a motívumra specifikusak, hanem a közvetlen közelben lévő határoló régiót, régiókat is jellemzik. Reményeink szerint a munkámban kapott eredmények beépítése egy TANGO vagy WALTZ szerű algoritmusba növelheti annak specifikusságát, vagyis tovább csökkentheti az esetleges fals pozitív találatok esélyét.

A munkámban kapott eredmények kiváló alapként szolgálnak abban a közeljövőben induló átfogó bioinformatikai vizsgálatban, melyben a rendezetlen fehérjeszakaszokon kialakuló fehérje – fehérje interakciós motívumok szerepét, szekvenciális sajátosságait, evolúciós kialakulásukat, változásait kívánjuk feltárni.

6. Köszönetnyilvánítás

Ezúton szeretném megköszönni témavezetőmnek, Kalmár Lajosnak segítségét, hogy segített nekem megismerni a bioinformatika tudományát, módszereit, lehetőségeit. Munkámat nagy odafigyeléssel irányítja, rendszeres konzultációink biztosítják a kutatás gördülékeny előrehaladását.

A dolgozatom elkészítésében témavezetőmnek és a csoportban dolgozó PhD hallgató Horváth Tamásnak tartozom köszönettel, Tamás nagy segítségemre volt az adatok részletes statisztikus feldolgozásának véghezvitelében.

Megköszönöm belső konzulensemnek, Gáspári Zoltánnak elsősorban az adminisztratív ügyekben való segítségnyújtást.

7. Melléklet

A vizuális értékelés alapján kiszűrt 219 tulajdonság aaindex azonosítója és rövid leírása kategóriákba rendezve:

Kategória	Kiválasztott aminosav indexek azonosítói és rövid leírásuk
Fizikai tulajdonságok	BIGC670101 (Residue volume), BULH740102 (Apparent partial specific volume), FAUJ880101 (Graph shape index), FAUJ880102 (Smoothed epsilon steric parameter), FAUJ880105 (STERIMOL minimum width of the side chain), LEVM760106 (van der Waals parameter R0)
Fizikokémiai tulajdonságok	BROC820101 (Retention coefficient in TFA), BROC820102 (Retention coefficient in HFBA), CHAM830107 (A parameter of charge transfer capability), CHAM830108 (A parameter of charge transfer donor capability), FASG760102 (Melting point), FASG760105 (pK-C), FAUJ880111 (Positive charge), FAUJ880112 (Negative charge), GUYH850101 (Partition energy), JOND750102 (pK (-COOH)), NOZY710101 (Transfer energy, organic solvent/water), OOBM770103 (Long range non-bonded energy per atom), OOBM850103 (Optimized transfer energy parameter), RACS770101 (Average reduced distance for C-alpha), RADA880101 (Transfer free energy from chx to wat), RADA880106 (Accessible surface area), RADA880108 (Mean polarity), ROBB790101 (Hydration free energy), VELV850101 (Electron-ion interaction potential), VHEG790101 (Transfer free energy to lipophilic phase), ZIMJ680102 (Bulkiness), COSI940101 (Electron-ion interaction potential values), GUOD860101 (Retention coefficient at pH 2), GUYH850102 (Apparent partition energies calculated from Wertz-Scheraga index), MIYS990101 (Relative partition energies derived by the Bethe approximation), MIYS990102 (Optimized relative partition energies - method A), MIYS990103 (Optimized relative partition energies - method B), MIYS990104 (Optimized relative partition energies - method C), MIYS990105 (Optimized relative partition energies - method D)
Flexibilitásra vonatkozó tulajdonságok	KARP850101 (Flexibility parameter for no rigid neighbors), VINM940102 (Normalized flexibility parameters (B-values) for each residue surrounded by none rigid neighbours), VINM940103 (Normalized flexibility parameters (B-values) for each residue surrounded by

	one rigid neighbours)
Hidrofóbicitásra vonatkozó tulajdonságok	BULH740101 (Transfer free energy to surface), CIDH920101 (Normalized hydrophobicity scales for alpha-proteins), CIDH920102 (Normalized hydrophobicity scales for beta-proteins), CIDH920103 (Normalized hydrophobicity scales for alpha+beta-proteins), CIDH920104 (Normalized hydrophobicity scales for alpha/beta-proteins), CIDH920105 (Normalized average hydrophobicity scales), EISD840101 (Consensus normalized hydrophobicity scale), FAUJ830101 (Hydrophobic parameter pi), GOLD730101 (Hydrophobicity factor), HOPT810101 (Hydrophilicity value), KYTJ820101 (Hydropathy index), LAW840101 (Transfer free energy, CHP/water), LEVM760101 (Hydrophobic parameter), MEIH800101 (Average reduced distance for C-alpha), MEIH800102 (Average reduced distance for side chain), MEIH800103 (Average side chain orientation angle), NAKH900104 (Normalized composition of mt-proteins), PARJ860101 (HPLC parameter), PLIV810101 (Partition coefficient), NADH010101 (Hydropathy scale based on self-information values in the two-state model), NADH010102 (Hydropathy scale based on self-information values in the two-state model), NADH010104 (Hydropathy scale based on self-information values in the two-state model), PONP930101 (Hydrophobicity scales), WILM950101 (Hydrophobicity coefficient in RP-HPLC, C18 with 0.1%TFA/MeCN/H2O), WILM950102 (Hydrophobicity coefficient in RP-HPLC, C8 with 0.1%TFA/MeCN/H2O), JURD980101 (Modified Kyte-Doolittle hydrophobicity scale), CASG920101 (Hydrophobicity scale from native protein structures), CORJ870101 (NNEIG index), CORJ870103 (PRIFT index), CORJ870104 (PRILS index), CORJ870105 (ALTFT index), CORJ870106 (ALTLS index), CORJ870107 (TOTFT index), CORJ870108 (TOTLS index)
Egyéb tulajdonságok	GRAR740101 (Composition), SNEP660101 (Principal component I), VENT840101 (Bitterness), WOLS870101 (Principal property value z1), PARS000101 (p-Values of mesophilic proteins based on the distributions of B values), ZHOH040102 (The relative stability scale extracted from mutation experiments), BAEK050101 (Linker index)
Statisztikai vizsgálatokból származó tulajdonságok	BIOV880101 (Information value for accessibility; average fraction 35%), BIOV880102 (Information value for accessibility; average fraction 23%), CHOC760102 (Residue accessible surface area in folded protein), FAUJ880110 (Number of full nonbonding orbitals),

	JANJ780102 (Percentage of buried residues), KRIW790101 (Side chain interaction parameter), KRIW790102 (Fraction of site occupied by water), MANP780101 (Average surrounding hydrophobicity), NAKH920105 (AA composition of MEM of single-spanning proteins), NAKH920108 (AA composition of MEM of multi-spanning proteins), NISK800101 (8 A contact number), NISK860101 (14 A contact number), PONP800101 (Surrounding hydrophobicity in folded form), PONP800102 (Average gain in surrounding hydrophobicity), PONP800103 (Average gain ratio in surrounding hydrophobicity), PONP800105 (Surrounding hydrophobicity in beta-sheet), PONP800106 (Surrounding hydrophobicity in turn), PONP800108 (Average number of surrounding residues), PRAM820101 (Intercept in regression analysis), RACS770102 (Average reduced distance for side chain), RACS820101 (Average relative fractional occurrence in A0(i)), RACS820108 (Average relative fractional occurrence in AR(i-1)), RACS820109 (Average relative fractional occurrence in AL(i-1)), RACS820110 (Average relative fractional occurrence in EL(i-1)), RACS820111 (Average relative fractional occurrence in E0(i-1)), RACS820113 (Value of theta(i)), RACS820114 (Value of theta(i-1)), RICJ880112 (Relative preference value at C3), RICJ880115 (Relative preference value at C-cap), RICJ880116 (Relative preference value at C'), ZIMJ680105 (RF rank), FUKS010103 (Surface composition of amino acids in extracellular proteins of mesophiles), FUKS010112 (Entire chain compositino of amino acids in nuclear proteins), YANJ020101 (Side-chain conformation by gaussian evolutionary method), BASU050101 (Interactivity scale obtained from the contact matrix), BASU050102 (Interactivity scale obtained by maximizing the mean of correlation coefficient over single-domain globular proteins), BASU050103 (Interactivity scale obtained by maximizing the mean of correlation coefficient over pairs of sequences sharing the TIM barrel fold)
Alkategóriába nem sorolható szerkezeti tulajdonságok	ANDN920101 (alpha-CH chemical shifts), CHAM830103 (The number of atoms in the side chain labelled 1+1), CHOP780207 (Normalized frequency of C-terminal non helical region), CHOP780210 (Normalized frequency of N-terminal non beta region), CHOP780211 (Normalized frequency of C-terminal non beta region), GEIM800108 (Aperiodic indices), GEIM800110 (Aperiodic indices for beta-proteins), GEIM800111 (Aperiodic indices for alpha/beta-proteins), OOBM850101 (Optimized beta-structure-coil equilibrium constant), PONP800107 (Accessibility reduction ratio), PTIO830101 (Helix-coil equilibrium constant), PTIO830102 (Beta-coil equilibrium constant),

	ROBB760113 (Information measure for loop), ROSG850101 (Mean area buried on transfer), ROSG850102 (Mean fractional area loss), SUEM840101 (Zimm-Bragg parameter s at 20 C), TANS770110 (Normalized frequency of chain reversal), VASM830101 (Relative population of conformational state A), VASM830103 (Relative population of conformational state E), YUTK870102 (Unfolding Gibbs energy in water, pH9.0), KOEP990101 (Structure-based conformational preferences of amino acids), KOEP990102 (Structure-based conformational preferences of amino acids), GEOR030108 (Linker propensity from helical)
Béta-redős szerkezeti elem kialakításában szerepet játszó tulajdonságok	CHOP780208 (Normalized frequency of N-terminal beta-sheet), CHOP780209 (Normalized frequency of C-terminal beta-sheet), CRAJ730102 (Normalized frequency of beta-sheet), GEIM800106 (Beta-strand indices for beta-proteins), GEIM800107 (Beta-strand indices for alpha/beta-proteins), ISOY800102 (Normalized relative frequency of extended structure), KANM800102 (Average relative probability of beta-sheet), KANM800104 (Average relative probability of inner beta-sheet), LEVM780102 (Normalized frequency of beta-sheet, with weights), LIFS790101 (Conformational preference for all beta-strands), LIFS790102 (Conformational preference for parallel beta-strands), LIFS790103 (Conformational preference for antiparallel beta-strands), NAGK730102 (Normalized frequency of beta-structure), PALJ810103 (Normalized frequency of beta-sheet from LG), PALJ810104 (Normalized frequency of beta-sheet from CF), PRAM900103 (Relative frequency in beta-sheet), QIAN880116 (Weights for beta-sheet at the window position of -4), QIAN880121 (Weights for beta-sheet at the window position of 1), QIAN880122 (Weights for beta-sheet at the window position of 2), QIAN880125 (Weights for beta-sheet at the window position of 5), QIAN880126 (Weights for beta-sheet at the window position of 6), ROBB760105 (Information measure for extended), ROBB760106 (Information measure for pleated-sheet), MUNV940103 (Free energy in beta-strand conformation), MUNV940104 (Free energy in beta-strand region), MUNV940105 (Free energy in beta-strand region), KIMC930101 (Thermodynamic beta sheet propensity)
Rendezetlen szerkezeti elem kialakításában szerepet játszó tulajdonságok	CHAM830101 (The Chou-Fasman parameter of the coil conformation), QIAN880130 (Weights for coil at the window position of -3), QIAN880131 (Weights for coil at the window position of -2), QIAN880132 (Weights for coil at the window position of -1), QIAN880133 (Weights for coil at the window position of 0), QIAN880134 (Weights for coil at the window position of 1), QIAN880136 (Weights for coil at the window position of 3), QIAN880139 (Weights for coil at the window position of 6),

	ROBB760112 (Information measure for coil)
Alfa-hélix szerkezeti elem kialakításában szerepet játszó tulajdonságok	BEGF750101 (Conformational parameter of inner helix), CHOP780201 (Normalized frequency of alpha-helix), CHOP780202 (Normalized frequency of alpha-helix), ISOY800101 (Normalized relative frequency of alpha-helix), KANM800101 (Average relative probability of helix), KANM800103 (Average relative probability of inner helix), QIAN880104 (Weights for alpha-helix at the window position of -3), QIAN880105 (Weights for alpha-helix at the window position of -2), QIAN880106 (Weights for alpha-helix at the window position of -1), QIAN880107 (Weights for alpha-helix at the window position of 0), ROBB760102 (Information measure for N-terminal helix), ROBB760103 (Information measure for middle helix), AURR980101 (Normalized positional residue frequency at helix termini N4'), AURR980102 (Normalized positional residue frequency at helix termini N'''), AURR980103 (Normalized positional residue frequency at helix termini N''), AURR980108 (Normalized positional residue frequency at helix termini N3), AURR980109 (Normalized positional residue frequency at helix termini N4), AURR980110 (Normalized positional residue frequency at helix termini N5), AURR980111 (Normalized positional residue frequency at helix termini C5), AURR980112 (Normalized positional residue frequency at helix termini C4), AURR980113 (Normalized positional residue frequency at helix termini C3), AURR980114 (Normalized positional residue frequency at helix termini C2), MUNV940102 (Free energy in alpha-helical region), BLAM930101 (Alpha helix propensity of position 44 in T4 lysozyme), FODM020101 (Propensity of amino acids within pi-helices)
Visszaforduló szerkezeti elem kialakításában szerepet játszó tulajdonságok	BEGF750103 (Conformational parameter of beta-turn), CHOP780101 (Normalized frequency of beta-turn), CHOP780203 (Normalized frequency of beta-turn), CHOP780212 (Frequency of the 1st residue in turn), CHOP780216 (Normalized frequency of the 2nd and 3rd residues in turn), CRAJ730103 (Normalized frequency of turn), ISOY800103 (Normalized relative frequency of bend), ISOY800104 (Normalized relative frequency of bend R), ISOY800105 (Normalized relative frequency of bend S), LEVM780103 (Normalized frequency of reverse turn, with weights), LEVM780106 (Normalized frequency of reverse turn, unweighted), LEWP710101 (Frequency of occurrence in beta-bends), PALJ810105 (Normalized frequency of turn from LG), PRAM900104 (Relative frequency in reverse-turn), ROBB760108 (Information measure for turn), ROBB760109 (Information measure for N-terminal turn), ROBB760110 (Information measure for middle turn), ROBB760111 (Information measure for C-terminal turn), MONM990101 (Turn propensity scale for transmembrane helices)

8. Irodalomjegyzék

1. Dobson C: **The structural basis of protein folding and its links with human disease.** *Philos Trans R Soc Lond B Biol Sci.* 2001, **356(1406)**:133-145.
2. Prusiner S: **Shattuck lecture--neurodegenerative diseases and prions.** *N Engl J Med.* 2001, **344(20)**:1516-1526.
3. Selkoe D, Podlisny M: **Deciphering the genetic basis of Alzheimer's disease.** *Annu Rev Genomics Hum Genet.* 2002, **3**:67-99.
4. Zoghbi H, Orr H: **Glutamine repeats and neurodegeneration.** *Annu Rev Neurosci.* 2000, **23**:217-247.
5. Jahn T, Radford S: **Folding versus aggregation: polypeptide conformations on competing pathways.** *Arch Biochem Biophys.* 2008, **469(1)**:100-117.
6. De B, Reumers J, Delgado B, Dopazo J, Schymkowitz J, Rousseau F: **An evolutionary trade-off between protein turnover rate and protein aggregation favors a higher aggregation propensity in fast degrading proteins.** *PLoS Comput Biol.* 2011, **7(6)**:10.
7. Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, Taddei N, Ramponi G, Dobson C, Stefani M: **Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases.** *Nature.* 2002, **416(6880)**:507-511.
8. Giurleo J, He X, Talaga D: **Beta-lactoglobulin assembles into amyloid through sequential aggregated intermediates.** *J Mol Biol.* 2008, **381(5)**:1332-1348.
9. Fändrich M, Fletcher M, Dobson C: **Amyloid fibrils from muscle myoglobin.** *Nature.* 2001, **410(6825)**:165-166.

10. Hutton M: **Missense and splice site mutations in tau associated with FTDP-17: multiple pathogenic mechanisms.** *Neurology*. 2001, **56(11 Suppl 4)**:S21-S25.
11. Booth D, Sunde M, Bellotti V, Robinson C, Hutchinson W, Fraser P, Hawkins P, Dobson C, Radford S, Blake C, Pepys M: **Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis.** *Nature*. 1997, **385(6619)**:787-793.
12. Riek R, Hornemann S, Wider G, Billeter M, Glockshuber R, Wüthrich K: **NMR structure of the mouse prion protein domain PrP(121-231).** *Nature*. 1996, **382(6587)**:180-182.
13. Näslund J, Haroutunian V, Mohs R, Davis K, Davies P, Greengard P, Buxbaum J: **Correlation between elevated levels of amyloid beta-peptide in the brain and cognitive decline.** *JAMA*. 2000, **283(12)**:1571-1577.
14. Sawaya M, Sambashivan S, Nelson R, Ivanova M, Sievers S, Apostol M, Thompson M, Balbirnie M, Wiltzius J, McFarlane H, Madsen A, Riek C, Eisenberg D: **Atomic structures of amyloid cross-beta spines reveal varied steric zippers.** *Nature*. 2007, **447(7143)**:453-457.
15. Conchillo-Solé O, de G, Avilés F, Vendrell J, Daura X, Ventura S: **AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides.** *BMC Bioinformatics*. 2007, **8**:65.
16. Fernandez-Escamilla A, Rousseau F, Schymkowitz J, Serrano L: **Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins.** *Nat Biotechnol*. 2004, **22(10)**:1302-1306.
17. Lacroix E, Viguera A, Serrano L: **Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters.** *J Mol Biol*. 1998, **284(1)**:173-191.
18. Maurer-Stroh S, Debulpaep M, Kuemmerer N, Lopez D, Martins I, Reumers J, Morris K, Copland A, Serpell L, Serrano L, Schymkowitz J, Rousseau F: **Exploring the sequence determinants of amyloid structure using position-specific scoring matrices.** *Nat Methods*. 2010, **7(3)**:237-242.

19. Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M: **AAindex: amino acid index database, progress report 2008**. *Nucleic Acids Res.* 2008, **36(Database issue)**:D202-D205.

20. Kawashima S, Ogata H, Kanehisa M: **AAindex: Amino Acid Index Database**. *Nucleic Acids Res.* 1999, **27(1)**:368-369.

9. Nyilatkozat

Alulírott Hatos András nyilatkozom, hogy a dolgozatomban felsorolt munkafolyamatokat saját magam végeztem el. Az egyes folyamatok tervezésében és eredmények értékelésében témavezetőm volt segítségemre. A statisztikai elemzésben Horváth Tamás volt segítségemre.

.....
Aláírás

Budapest, 2013. január 9.