

Pázmány Péter Katolikus Egyetem

Roska Tamás Műszaki és
Természettudományi Doktori Iskola



Szabó András László

Posztzinaptikus fehérjehálózatokban található fehérjék
fázisszeparációjának vizsgálata számításos és kísérleti
megközelítések kombinációjával

PhD disszertáció tézisei

Témavezető:

Dr. Gáspári Zoltán

2025

Célkitűzések

A fő célja az itt bemutatott kutatómunkának, hogy megvizsgálja azon összefüggéseket, amelyek teoretikusan fennállnak a posztszinaptikus fehérjehálózatokat formáló fázisszeparáció jelensége, a GKAP és LC8 fehérjék komplexformálása, illetve a magányos α -hélixek (single α -helix, SAH) és egyéb töltött motívumok között. Ez utóbbi összefüggés kivizsgálásához szükséges a fehérjeszekvenciákban található töltött motívumok proteom-szintű tanulmányozása, ami magába foglalja:

- Egy proteom-szintű adathalmaz létrehozását, amely tartalmazza a vizsgálni kívánt motívumokat, valamint a fázisszeparációért felelős régiókat.
- Az adathalmaz redundanciájának minimalizálását.
- A vizsgált motívumok és a fázisszeparáció jelensége közötti lehetséges összefüggések kiértékelését.

A GKAP és LC8 fehérjék közötti komplex kialakulása megfigyelhető az oldatban lévő részecskék méreteloszlásának mérésével. Ezt el lehet érni egy diffúzió alapuló kísérleti eljárással, amely fluoreszcens mikroszkópiai módszereket kombinál mikrofluidikai eszközökkel. Egy ilyen eljárás kifejlesztése a következő lépéseket vonja maga után:

- A kísérleti eljárás megtervezése.
- Megfelelő kialakítású mikrofluidikai eszközök legyártása.
- Fluoreszcens minták előkészítése kalibrációhoz és elemzéshez.
- Azon elemző szoftver kifejlesztése, amely kiértékeli a mért adatokat.

Módszerek

A proteom-szintű adathalmaz létrehozása

Az általam használt referencia proteom 20659 emberi gént foglalt magába, génenként egy izoformával, amelyeket a manuálisan összeállított SwissProt adatbázisból gyűjtöttem ki. Az így kapott adathalmazt a PhaSepDB

adatbázisban annotált, fázisszeeparációért felelős régiókkal egészítettem ki, amelyek szerepe kísérletesen igazolt. Az adathalmazt tovább bővítő, ismétlődő töltésmintázatokkal rendelkező motívumokat (charged residue repeat, CRR) az FT_CHARGE algoritmussal detektáltam, amely a szekvencia töltés korrelációs függvényének Fourier-transzformáltját kutatja át szabályosan ismétlődő, töltött aminosavak után. Ezen motívumok közül kiemeltem a SAH-okat, melyek karakterisztikus frekvenciája a Fourier-térben $1/9$ -tól és $1/6$ -ig terjed. Végül, az adathalmazt tovább bővítettem olyan régiókkal, amelyek vagy nagy arányban tartalmaztak töltött aminosavakat, vagy az össztöltésük kiugró volt pozitív vagy negatív irányban. Ezen régiók felmérésénél nem volt feltétel, hogy a hozzájuk tartozó töltött aminosavaknak bármilyen mintázata legyen. Két különböző módszerrel mértem fel ezeket az ún. töltésben sűrű régiókat (charge-dense region, CDR). Az egyik módszer az alapján azonosította ezen motívumokat, ha a régióban lévő töltött aminosavak aránya elérte egy küszöbértéket. A másik módszer a régiók össztöltését hasonlította egy semleges érték körüli sávhoz. A két különböző módszerrel kapott motívumokat ezért „előjeles” (signed) és „előjel nélküli” (unsigned) CDR-eknek neveztam el, és mostantól sCDR és uCDR néven fogok utalni rájuk.

Az adathalmaz redundanciájának minimalizálása

A fenti módon elkészített referencia proteomot a szekvenciák közötti hasonlóság alapján csoportosítottam a CD-HIT program segítségével, 0,9, 0,7, valamint 0,5 küszöbértékek mellett. Minden szekvenciát a legjobb csoporthoz rendelte hozzá a program azok közül, amelyek elérték a küszöbértéket.

A vizsgált motívumok és a fázisszeparáció jelensége közötti lehetséges összefüggések kiértékelése

A referencia proteomban, valamint annak csoportosított variánsaiban található elemeket két tulajdonság alapján rendeztem 2x2-es táblázatokba. Az egyik tulajdonság töltött motívumok jelenléte – avagy hiánya – volt. A másik tulajdonság a fázisszeparációra való hajlam volt, amit az ahhoz kapcsolódó – PhaSepDB-ben annotált – régiók jelenléte alapján határoztam meg. A két tulajdonság közötti összefüggés mértékével arányos P-értékeket Fisher tesztek (Fisher's exact test of independence) segítségével számoltam ki. Továbbá, a különböző töltött motívumok jelenlétét kiértékeltem a fázisszeparációban való részvétel indikátoraként, ROC (receiver operating characteristic) analízissel. Ehhez a referencia proteomot a motívumok detektálásakor kapott pontszámok alapján átrendeztem, amit követően a valós pozitív, valós negatív, fals pozitív és fals negatív rátákat a fázisszeparációban való részvétel alapján megállapítottam. A valós pozitív elemeket egyenként tovább vizsgáltam.

A kísérleti eljárás megtervezése

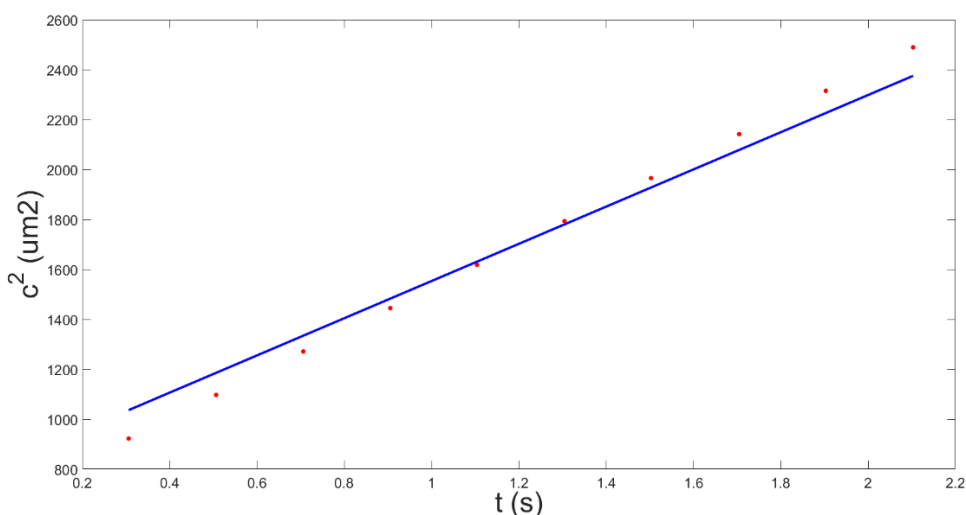
A GKAP és LC8 fehérjét, valamint a komplexüket a hidrodinamikai átmérőik alapján különböztettem meg, a Stokes-Einstein egyenletnek megfelelően, ami csak lamináris áramlásra jellemző alacsony Reynolds számok mellett teljesül:

$$1. \text{ Egyenlet: } D = \frac{k_B T}{6\pi\eta r}$$

A részecskéket egy jól meghatározott kezdőpozícióba helyeztem egy mikrofluidikai fókuszáló eszköz használatával, amelyben az oldalsó bemeneteken beáramló pufferoldatok szorították egy szűk sávba a középső bemeneten bejuttatott mintát. Ezt követően a minta és a pufferoldatok lamináris áramlás révén haladtak tovább a lapos csatornában, amely mentén előre meghatározott helyen készítettem felvételeket a minta áramlásra

merőleges eloszlásáról, ami így csak diffúzió révén változhatott. A részecskék eloszlását követő lineárisan növekvő függvény alapján megbecsültem a részecskék diffúziós koefficiensét:

2. Egyenlet: $D = \frac{c^2}{4t}$



1. ábra: Az oldott részecskék eloszlásának mértéke a csatorna különböző pontjain, amiket az ezen pontokban felvett fluoreszcens intenzitás profilokra illesztett Gauss-görbék varianciájával (c^2) becsültem meg. Az x-tengely azt írja le, hogy mennyi idő telik el, amíg egy adott részecske a csatorna elejétől a különböző mérési pontokhoz ér. A minta ennél a mérésnél zöld fluoreszcens fehérje (enhanced green fluorescent protein, EGFP) volt. [2, 1. ábra]

Eszközök legyártása és minták előkészítése

A mikrofluidikai eszközöket Laki Mária készítette el. A kialakításuk iteratív módon fejlődött az általam végzett tesztek alapján. Az eszközök belső felületét bovine serum albumin (BSA) oldattal kezeltem minden mérést megelőzően, hogy a mérés során minél lassabban tapadjanak ki arra fluoreszcens részecskék. Az EGFP és fluoreszcens mikrogöngy mintákat Jáger Edit Andrea készítette elő. A GKAP és LC8 konstrukciókat, illetve a

GKAP és LC8 által képzett komplexeket tartalmazó mintákat Nagy-Kanta Eszter expresszáltatta, tisztította, illetve jelölte meg fluoreszcens festékekkel. Minden fehérje mintát fluorescein (FITC) festékekkel jelöltünk meg, amelyet több másik fluoreszcens festék tesztelése után választottam ki. A mérési eljárást FITC-jelölt Drebrin mintán is teszteltem, amit Varga Soma szolgáltatott.

Az elemző szoftver fejlesztése

A mért adatokat egy MATLAB környezetben írt elemző szoftveren keresztül dolgoztam fel, ami Excel fájlokat igényel bemenetként. Ezen fájloknak tartalmaznia kell a mérési pontokon felvett fluoreszcens és világos látóterű intenzitás profilokat, azok pontos pozícióját, illetve olyan általános tényezőket, mint az abszolút hőmérséklet és az expozíciós idő. A részecskék mérési pontok közötti áthaladásának idejét a beállított áramlási sebesség alapján becsültem meg. A fluoreszcens intenzitás profilokat a mérésből származó műtermékek eltávolítása után normalizáltam, hozzá igazítottam a tengelyekhez őket, és Gauss-függvényeket illesztettem rájuk. Egy profilhoz két függvény lineáris kombinációját illesztettem, amely két függvényt két külön sorozathoz rendeltem hozzá. A sorozatok mérési pontonként egy-egy függvényt tartalmaztak, amelyek varianciáit az adott mérési pontokhoz tartozó időkomponensekkel párosítottam, az 1. ábrának megfelelően. Az így kapott, mérésenként kettő darab, lineáris függvény meredeksége arányos a méréshez használt mintában lévő részecskék diffúziós koefficienseivel.

Eredmények

PhD tanulmányaim alatt összetett kutatási projekteket folytattam fehérjék fázisszeparációjának feltárása végett, a poszt-szinaptikus fehérjehálózat kontextusában. Eme projektek több eredményhez és konklúzióhoz vezettek, melyek közül a legjelentősebbeket az alábbi tézispontokban összegzem:

1. Tézispont: Meglévő számítási módszereket alkalmazva, valamint új módszereket kifejlesztve azonosítottam fehérje szekvenciákban található töltött motívumok több típusát. [1]

a) **Módszerek:** Összeállítottam egy referencia proteomot a SwissProt adatbázisban található emberi fehérje szekvenciákból. Kiegészítettem eme adathalmazt olyan motívumok pozícióival, amelyek szerepe a fázisszeparáció jelenségében kísérletileg bizonyított, a PhaSepDB2.0 adatbázis annotációi szerint. Kibővítettem az adathalmazt a(z) FT_CHARGE algoritmus által meghatározott CRR szegmensekkel, külön kiemelve a karakterisztikus töltésmintázatuk alapján azonosított SAH motívumokat. Tovább bővítettem az adatsort magas töltöttségű régiókkal, amelyeket a saját algoritmusaimmal azonosítottam. Eme algoritmusok ablakoló függvényeken és két eltérő pontozási rendszeren alapultak, a belőlük származó, CDR megnevezésű szakaszok így két csoportra lettek bontva. Minimalizáltam a teljes referencia proteom redundanciáját a CD-HIT szoftver segítségével (1. táblázat).

b) **Eredmények:** Az összeállított adathalmaz vizsgálatával kimutattam, hogy a benne található fehérjék jelentős része tartalmaz legalább egy CDR szakaszt. A transzmembrán szegmensekkel rendelkező fehérjéket kihagytam a további vizsgálatokból, mivel a fázisszeparációban résztvevő fehérjék várhatóan oldott állapotban vannak jelen.

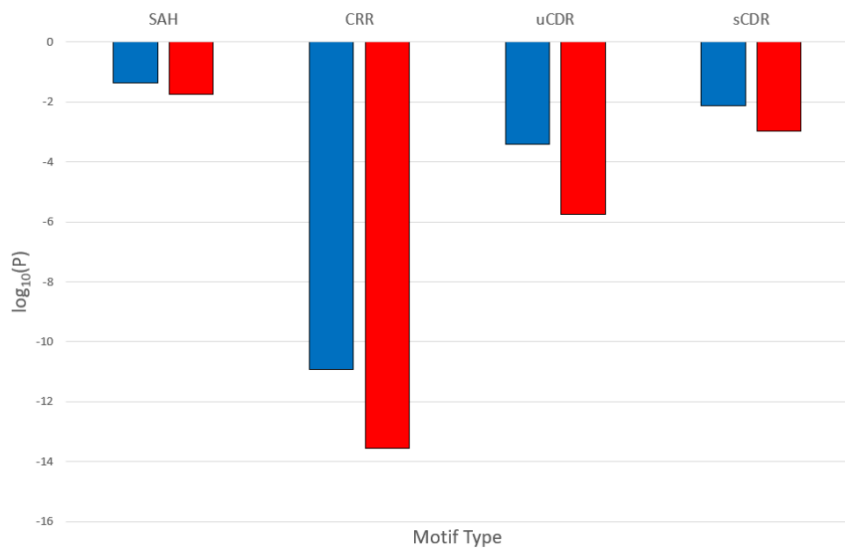
Szekvenciában található motívumok	Teljes proteom	Csökkentett redundanciával rendelkező proteomok		
		90%	70%	50%
Fehérjék száma	20 659	19 638	18 294	15 672
uCDR szakaszok	9 731	9 314	8 792	7 669
sCDR szakaszok	14 065	13 471	12 757	11 097
CRR szakaszok	1 054	1 025	985	910
SAH szakaszok	134	131	126	118

1. táblázat: Azon szekvenciák száma a referencia proteomban, valamint annak csökkentett redundanciával rendelkező variánsaiban, amelyek tartalmaznak legalább egy fajta töltött motívumot.

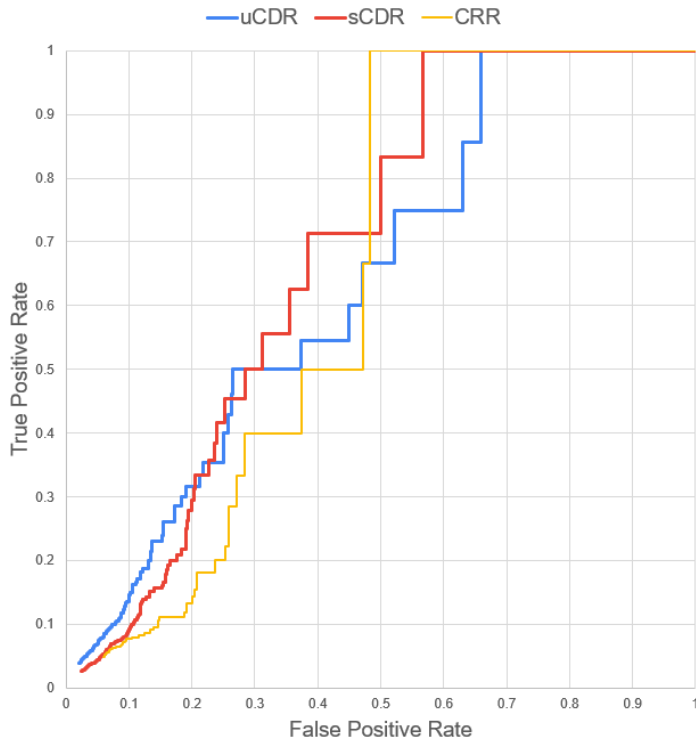
2. Tézispont: Bizonyítottam robusztus összefüggések fennállását a különböző töltött motívumok és az azokat tartalmazó fehérjék fázisszeparációra való hajlamossága között. Elsősorban negatív összefüggéseket találtam, vagyis a vizsgált motívumok hiánya teszi valószínűtlenné a jelenségben való részvételt. [1]

a) **Módszerek:** 2x2 táblázatokba rendeztem a referencia proteom, valamint annak csökkentett redundanciájú variánsainak elemeit. A táblázatokon Fisher tesztek végeztem el, amelyek kimutatják, hogy egy fehérje fázisszeparációban való részvétele, illetve a szekvenciájában található motívumok jelenléte egymástól független változók-e. A tesztek megismételttem véletlenszerűen összeállított szekvenciák halmazain, melyek méreteloszlása és aminosav összetétele megegyezett a referencia proteommal. ROC analízis segítségével vizsgáltam meg, hogy a különböző töltött motívumok jelenléte használható-e az őket tartalmazó fehérjék fázisszeparációban való részvételének indikátoraként.

- b) **Eredmények:** Az elvégzett Fisher tesztekkel kimutattam, hogy a fázisszeparációban részt vevő fehérjék CRR szakaszokban gazdagok (2. ábra). Ezen összefüggés SAH szakaszok esetében gyengébb, de továbbra is szignifikáns. A véletlenszerűen összeállított adatsorok ezekkel egyező eredményeket szolgáltatnak, bizonyítva a feltárt összefüggések robusztusságát.



2. ábra: A Fisher tesztekől származó P-értékek tízes alapú logaritmusai, melyek megmutatják az egyes töltött motívumok jelenléte és a fehérjék fázisszeparációban való részvétele között fennálló összefüggés mértékét. Ezen vizsgálatokat megismétltem a transzmembrán szegmenseket tartalmazó fehérjék kizárásával (kék). [1, 3. ábra]



3. ábra: A vizsgált motívumok fázisszeparációt illető prediktív képességének kimutatása ROC analízis segítségével. [1, 6. ábra]

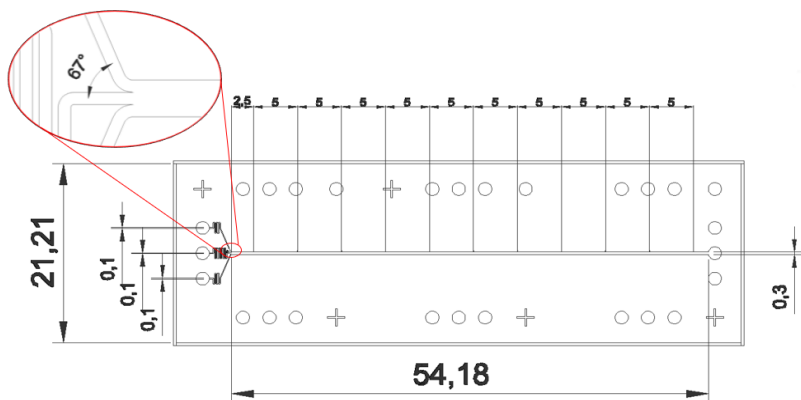
Habár a töltött motívumok jelenléte önmagában a fázisszeparáció gyenge indikátorának bizonyult (3. ábra), a Fisher tesztek eredményét figyelembe véve hiányuk erős indikátorként szolgál amellet, hogy az adott fehérje nem vesz részt a jelenségben.

3. Tézispont: Kifejlesztettem egy kísérleti eljárást, amely alkalmas fehérjék és azok komplexeinek méretének meghatározására, alapul véve a részecskék lamináris áramlása közben végbemenő laterális diffúziót. [2]

Módszerfejlesztés: Megállapítottam, hogy a fluorescein kellő jel-zaj viszony mellett festi meg a GKAP, LC8 és Drebrin fehérjék adott konstrukcióit anélkül, hogy megzavarná az első kettő komplex formálását. Részt vettem egy mikrofluidikai eszköz

megtervezésében, amely képes fluoreszcensen jelölt minták fókuszálására, valamint lamináris áramoltatására. Optimalizáltam az eszközben áramló minták által emittált fluoreszcens jel felvételét. Kifejlesztettem egy elemző szoftvert, amivel a felvett jeleket a mintában lévő részecskék hidrodinamikai átmérőjévé lehet konvertálni.

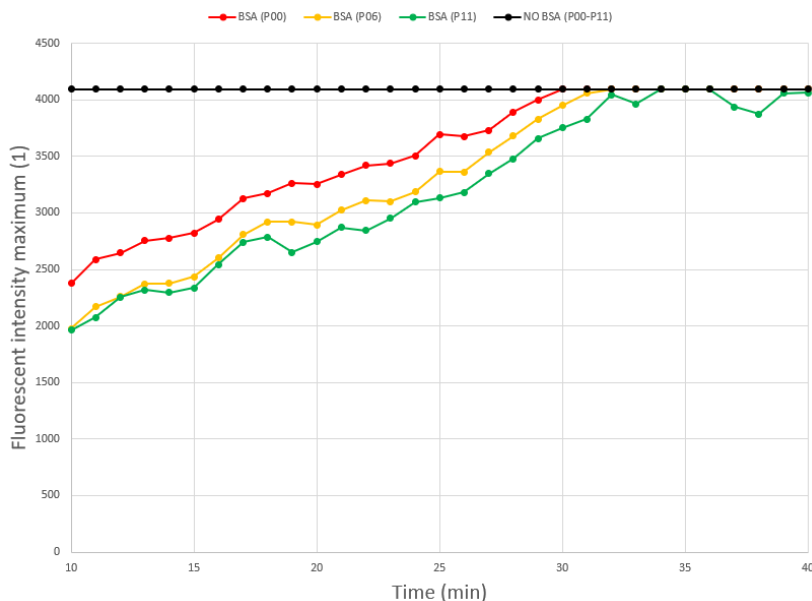
Megállapítottam, hogy egy tárgylemezre legfeljebb három cella hosszúságú mikrofluidikai eszközt lehet ráilleszteni. Emiatt a méréshez használt csatorna hossza 54 mm volt a véglegesített eszközön, mivel egy kanyarulatokkal meghosszabbított csatorna megtörné a minta lamináris áramlását (4. ábra).



4. ábra: A végleges eszköz kialakítása tartalmaz ekvidisztáns jelölőket, amik elősegítik a fluoreszcens és világos látóterű jelek felvételét az előre meghatározott mérési pontokban. [2, 7. ábra]

Egy másik megállapításom, hogy az eszközök belső felületét *bovine serum albumin* (BSA) oldattal kell kezelni a mérést megelőzően, ami a fluoreszcens részecskék kitapadásának lelassítását szolgálja. BSA kezelés nélkül elég fluoreszcens részecske tapadna ki ahhoz 10 percen belül, hogy azok együttes intenzitása meghaladja a mikroszkóp felső határértékét (5. ábra).

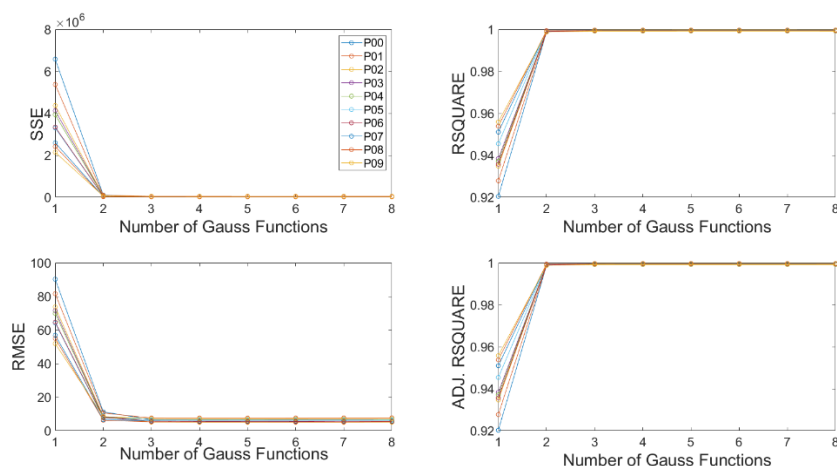
Legalább 10 percnak kell eltelnie a minta megjelenését követően, mire beáll az eszközön belüli steady state.



5. ábra: Fluoreszcens csúcsintenzitás a mikrofluidikai eszköz bemeneteinek találkozásánál (P00), az azt követő csatorna felénél (P06), illetve a csatorna végénél (P11). Az intenzitás percenként volt megmérve fél órán keresztül két különböző mikrofluidikai eszközben, melyek közül csak az egyiket kezeltam BSA oldattal. A minta 0,05 μm átmérőjű fluoreszcens mikrogöngyökből állt mindkét esetben. Az első mérést 10 perccel az után végeztem el, hogy a minta beért az eszközbe, amely várakozási idő szükséges a steady state beállításához. [2, 8. ábra]

Továbbá, az elemző szoftvert úgy állítottam be, hogy két Gauss-függvény lineáris kombinációjaként közelítse a mérésből származó fluoreszcens intenzitás profilokat. Erre a döntése az alábbi ábrán látható „illesztés jóságát” leíró metrikák vezettek rá,

amelyek lényegesen javulnak két függvény lineáris kombinációjának használatával, de nem javulnak tovább jelentős mértékben több, mint két függvény használata esetén (6. ábra).



6. ábra: Az illesztés jóságát leíró (goodness of fit, GoF) metrikák, különböző mérési pontokra (P00-P09) kiszámolva nyolc darab, komplexitásban eltérő modell mellett. Az itt bemutatott mérés során EGFP volt a minta, de más mintákkal megismételve ugyanezen tendenciák ismétlődtek mind a négy GoF metrika esetében. [2, 9. ábra]

4. Tézispont: Bizonyítottam, hogy a mérési eljárás eredményei konzisztensek az előzetes ismeretekből származó adatokkal, valamint, hogy a kapott eredmények precízebbek a dinamikus fényszórásnál 10 nm alatti részecskeátmérő esetén. [2]

- a) **Módszerek:** Méréseket végeztem a véglegesített eljárással fluoreszcensen jelölt Drebrin (D233) és GKAP (GKAP-PBM és GKAP-DLC2) konstrukciókon, valamint GKAP-DLC2 + LC8 komplexeken, ahol a GKAP-DLC2 monomereket jelöltük összekeverés előtt. Ezekon felül EGFP mintákon, illetve 50, 200

és 1100 nm névleges átmérőjű fluoreszcens mikrogöngyökön is végeztem méréseket.

- b) **Eredmények:** Az eljárással becsült sugarak átlagértéke konzisztens 10 nm átmérő alatti részecskék esetén. Ezen a mérettartományon a szórás alacsonyabb a diffúzió alapuló eljárás esetében, mint dinamikus fényszórás alkalmazásával (2. táblázat). Nagyobb részecskénél az eljárás pontossága és precizitása is romlik.

Minta	Várt sugár (nm)	Becsült sugár (nm)
GKAP-PBM	1,39	$1,16 \pm 0,10$
GKAP-DLC2	$1,27 \pm 0,72$ $1,90 \pm 1,00$	$1,36 \pm 0,30$
GKAP-DLC2 + LC8	$2,17 \pm 0,44$ $3,11 \pm 0,76$	$1,64 \pm 0,36$
D233	1,81	$1,45 \pm 0,39$
EGFP	$2,72 \pm 1,01$	$1,72 \pm 0,42$
50 nm gyöngy	$22,68 \pm 2,11$	$9,28 \pm 11,09$
200 nm gyöngy	$97,09 \pm 9,51$	$15,15 \pm 13,36$
1100 nm gyöngy	$461,61 \pm 35,59$	$6,62 \pm 5,78$

2. táblázat: A megmért minták listája, azok dinamikus fényszórás alapján (vagy, a GKAP-PBM és D233 esetében, molekulatömeg alapján) várt sugara, illetve a diffúziós eljárással becsült sugaruk. A fluoreszcensen jelölt GKAP-DLC2 konstrukció és GKAP-DLC2 + LC8 komplex esetében két dinamikus fényszórás kísérletet végeztünk el, amelyeket külön sorban tüntettem fel. A becsült sugarak átlagértékét és szórását több mérés adataiból számoltam ki.

Eredményeim előrelépést jelentenek a fázisszeparációról, illetve a poszt-szinaptikus fehérjehálózatról szóló ismereteinkben. Tanulmányaim során több, ezen tudományos területek művelésére alkalmas számítási és kísérleti módszert fejlesztettem ki. Számos robusztus összefüggést fedeztem fel töltött motívumok és a fázisszeparáció jelensége között. Kifejlesztettem egy a poszt-szinaptikus fehérjehálózatok elemeinek vizsgálatára alkalmas mérési eljárást, amely betekintést nyújt az ezen komplex hálózatok belső szerkezetét mozgató interakciókba.

Publikációk

1. A. L. Szabó, A. Sánta, R. Pancsa, Z. Gáspári, "Charged sequence motifs increase the propensity towards liquid-liquid phase separation," *FEBS Lett.*, vol. 596, no. 8, pp. 1013-1028, Apr 2022
2. A. L. Szabó, E. Nagy-Kanta, S. Varga, E. A. Jáger, C. I. Pongor, M. Laki, A. J. Laki, Z. Gáspári, "Diffusion-based size determination of solute particles: a method adapted for PSD proteins," *FEBS Open Bio*, Accepted for publication, preprint doi:10.1101/2025.02.05.636588
3. Z. Harmat, A. L. Szabó, O. Tőke, Z. Gáspári, "Different modes of barrel opening suggest a complex pathway of ligand binding in human gastrotropin," *PLoS ONE*, vol. 14, no. 5, e0216142, May 2019, doi:10.1371/journal.pone.0216142

Konferencia előadások

4. A. L. Szabó. (Jul 2022). Charged sequence motifs increase propensity towards liquid-liquid phase separation. Presented at IUBMB-FEBS-PABMB Young Scientists' Forum 2022, Vimeiro, Portugal. [Poster]. Available: Young Scientists' Forum 22 Programme and Abstract Book (pp. 163)
5. A. L. Szabó., A. Sánta, R. Pancsa, Z. Gáspári. (Jul 2022). Charged sequence motifs increase propensity towards liquid-liquid phase

separation. Presented at The Biochemical Global Summit 2022, Lisbon, Portugal. [Poster]. Available:

<https://febs.onlinelibrary.wiley.com/doi/epdf/10.1002/2211-5463.13440> (pp. 321)

6. A. L. Szabó, A. Sánta, R. Pancsa, E. A. Jáger, C. Pongor, A. J. Laki, Z. Gáspári. (Nov 2022). Phase separation of postsynaptic proteins: insights from bioinformatics and microfluidics. Presented at 1st FEBS-IUBMB-ENABLE Conference, Seville, Spain. [Poster].
7. A. L. Szabó, E. A. Jáger, C. I. Pongor, A. J. Laki, Z. Gáspári. (Jun 2024). Diffusion-based analysis of phase separating PSD proteins: combining microfluidics with fluorescent microscopy techniques. Presented at 48th FEBS Congress, Milano, Italy. [Poster]. Available: <https://febs.onlinelibrary.wiley.com/doi/epdf/10.1002/2211-5463.13837> (pp. 135)

Hozzájárulásaim

A fenti folyóiratcikkekbe sűrített tanulmányok számos kutató kemény munkáját és hozzáértését igényelték, amiért én végtelenül hálás vagyok. Az itt felsorolt feladatok foglalják össze az én hozzájárulásomat a közös munkához: Összeállítottam egy proteom-szintű adathalmazt nyílt adatbázisokból, valamint SAH és egyéb CRR szakaszokat tartalmazó fájlokból. A CD-HIT programmal csoportosítottam az adathalmazt, valamint MATLAB kódokat írtam Fisher tesztek, ROC analízisek, illetve egyéb, szekvencia összetételt és GO kifejezéseket illető vizsgálatok elvégzéséhez. Hozzájárultam a töltött motívumokat tartalmazó és fázisszeparációban is résztvevő fehérjék tanulmányozásához. Részt vettem a diffúzió alapuló mérési eljárás, illetve az ahhoz szükséges mikrofluidikai csatornák megtervezésében, valamint a mérések elvégzésében. Kifejlesztettem egy elemző szoftvert a mért adatok feldolgozásához. Részt vettem az összes eredmény kiértékelésében és értelmezésében.