

Pázmány Péter Katolikus Egyetem
Roska Tamás Műszaki és
Természettudományi Doktori Iskola



Posztzinaptikus fehérjekomplexek számítógépes modellezése

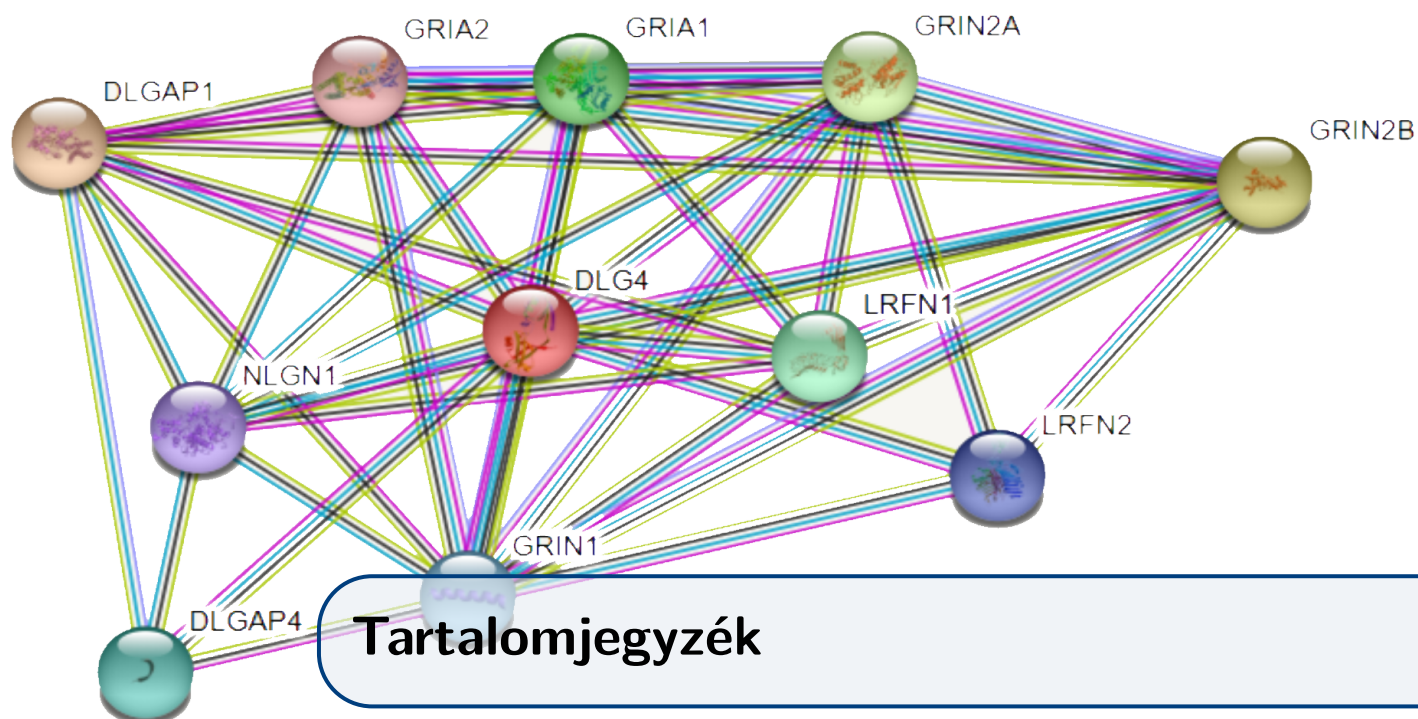
PhD Disszertáció

Miski Marcell

Témavezető: Prof. Dr. Csikász-Nagy Attila DSc

2024

PÁZMÁNY PÉTER KATOLIKUS EGYETEM,
Roska Tamás Műszaki és Természettudományi Doktori Iskola
A *PhD* fokozat megszerzésére benyújtott dolgozat
2024. július 25.

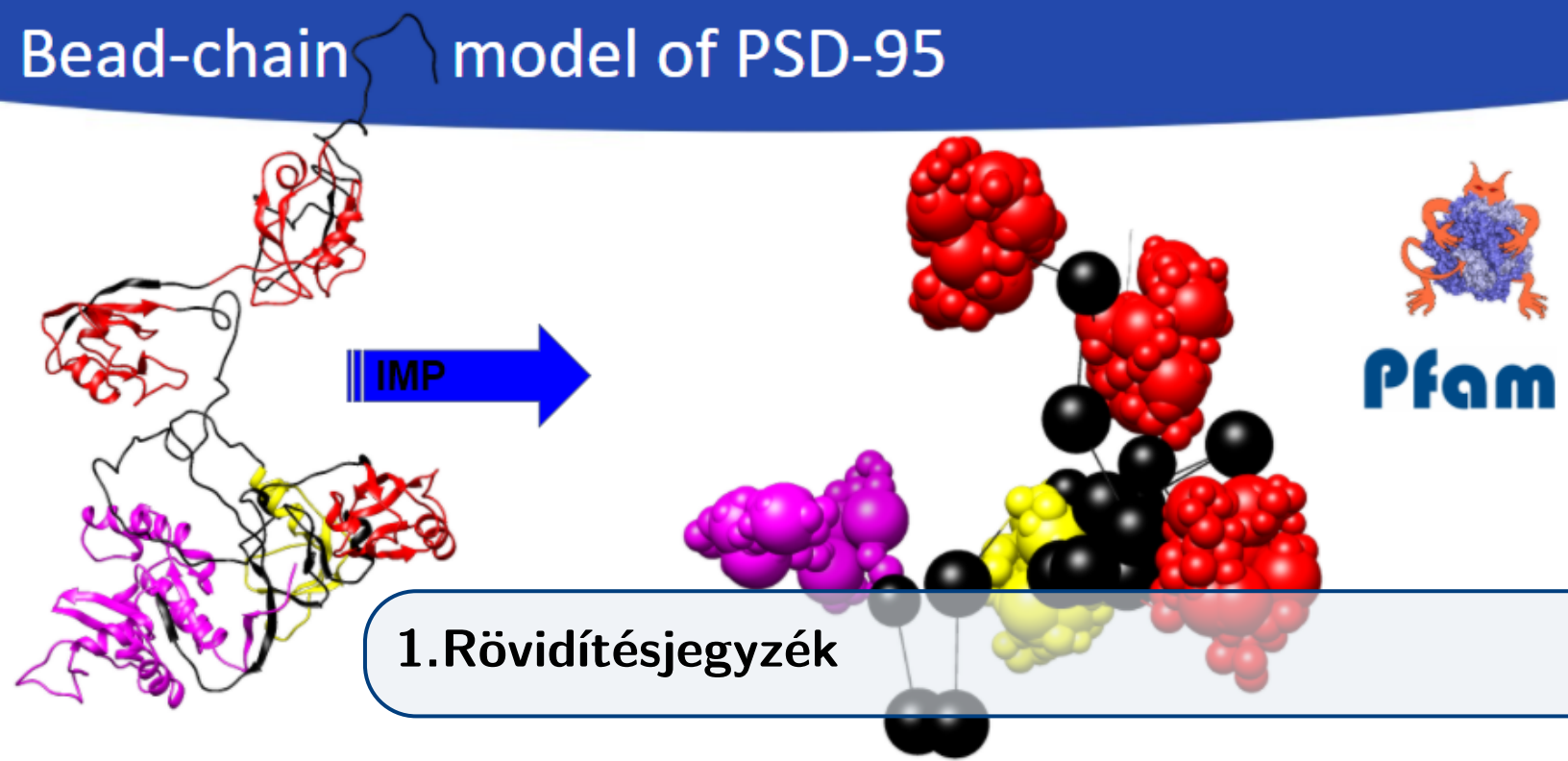


1	Rövidítésjegyzék	6
2	Kivonat	9
3	Abstract	11
4	Bevezetés	13
4.1	A posztszinaptikus denzitás	13
4.2	A szinaptom elmélet	15
4.2.1	Fehérjegyakoriságok a posztszinaptikus denzitásban	16
4.3	A posztszinaptikus denzitást vizsgálatára használt kísérletes módszerek	16
4.3.1	Fehérjeösszetétel vizsgálatára alkalmas módszerek	16
4.3.2	Szuperrezolúciós mikroszkópai eljárások	17
4.3.3	PALM és STORM	20
4.3.4	Génkiütéses viselkedésvizsgálatok	21
4.3.5	Fehérjék térszerkezetének meghatározása	21
4.4	Fehérjekomplexek kialakulásának ágens alapú szimulációja	22
4.5	Fehérjekomplexek 3D szerkezetének modellezése	23
4.6	A posztszinaptikus denzitás fehérjei	24
4.6.1	A Kir2.1 káliumion-csatorna	25
4.6.2	Az NMDA receptor	25
4.6.3	Az AMPA receptor	26
4.6.4	A PSD-95 fehérje	27
4.6.5	A SynGAP fehérje	28

4.6.6	A GKAP fehérje	28
4.6.7	A Shank3 fehérje	29
4.6.8	A Shank1 fehérje	29
4.6.9	A Homer1 fehérje	30
5	Célkitűzések	31
6	Tézisek és közlemények	33
7	Módszerek	35
7.1	Fehérjeadatbázisok	35
7.1.1	UniProt	35
7.1.2	DisProt	37
7.1.3	Pfam	37
7.1.4	Gene Ontology	39
7.1.5	RCSB - Protein Data Bank (PDB)	39
7.1.6	Cath	40
7.1.7	BrainSpan Atlas	41
7.1.8	Postsynaptic Interaction Database - PSINDB	41
7.2	Fehérje- és kötésstabilitást jósló módszerek	41
7.2.1	CAD-pontszám	42
7.2.2	NeEMO	43
7.3	A Gillespie algoritmus, amelyet a Cytocast szoftver is használ	44
7.4	K-means algoritmus	45
7.5	Főkomponens Analízis	47
7.6	t-próba	47
7.7	A Shapiro-Wilk Teszt Alkalmazása és Jelentősége	50
7.7.1	A Shapiro-Wilk Teszt Alapjai	51
7.7.2	Alkalmazási Területek és Előnyök	51
7.7.3	Kritikus Értékelés és Limitációk	51
7.8	A Szignifikancia Szint Korrigálása: Bonferroni és Benjamini-Hochberg Módszerek	51
7.8.1	Bonferroni korrekció	51
7.8.2	Benjamini-Hochberg korrekció	52
7.8.3	Érvek a Benjamini-Hochberg módszer mellett	52
7.9	Háromdimenziós modellekhez használható szerkezetek	52
7.10	A szuperfelbontású mikroszkópia a szimulációk és modellek validálására	55
8	Eredmények	56
8.1	Az ágens-alapú szimuláció bemeneti modelljének kidolgozása	56

8.2	A posztzinaptikus denzitás mennyiségi és minőségi modellezése	57
8.2.1	Fehérjekomplexek integratív modellezése	57
8.3	A szimulációk reprodukálhatósága	59
8.3.1	A szimulációs idő hatása a komplexek méretére	61
8.4	A mutációk hatásainak kiszámítása kötéseerősség predikcióval	63
8.4.1	A Shank1 R743H mutáció kiválasztása és modellezése	64
8.5	Az ágens-alapú szimulációk értékeléséhez kidolgozott eljárások	64
8.5.1	A keletkezett komplexek gráfszerkezetének vizsgálata	64
8.5.2	A komplexek azonosítása	65
8.5.3	A szimulációk hozzáadott információtartalmára utaló mérőszám meghatározása	68
8.5.4	A legmagasabb információtartalmú komplex azonosítása	69
8.5.5	Fehérjekomplex abundanciák összehasonlítása	70
8.6	A szimulációk hozzáadott információtartalommal rendelkeztek	71
8.6.1	Az AMPAR/PSD-95/SYNGAP komplex a leginformatívabb	72
8.6.2	A régiótípusok számottevően nem különböztek el egymástól	72
8.6.3	A komplexek eloszlásának alapvető jellemzőit elsősorban a fehérjegyakoriság határozza meg	74
8.7	Különböző fehérjekapcsolatok hatásai a komplexekre	77
8.7.1	Homer1 tetramerizáció és Shank3 szelfasszociáció hatásai	77
8.7.2	A hasonló fehérjegyakoriságok ritkán, de eredményezhetnek eltérő komplexeloszlást	80
8.7.3	A Shank:GKAP interakció gyengítése korlátozott változásokat okoz egyes jól meghatározott régiókban	83
8.8	A felépített háromdimenziós modellek	85
9	Diszkusszió	89
9.1	Megközelítésem indoklása	89
9.1.1	A fehérjegyakorisági adatsorok eloszlásának kritikai áttekintése	89
9.2	A PSD-komplexek modelljei – milyen messze vagyunk a valóságtól?	90
9.3	Összefoglalás és tézispontok	91
9.4	Kutatásom kilátásai: lehetséges hosszútávú irányai és kibontakozása	94
10	Köszönetnyilvánítás	95
11	Irodalomjegyzék	96
12	Kiegészítő ábrák	118
13	Kiegészítő Táblázatok	123
14	Kiegészítő Python kódok	126

Bead-chain model of PSD-95

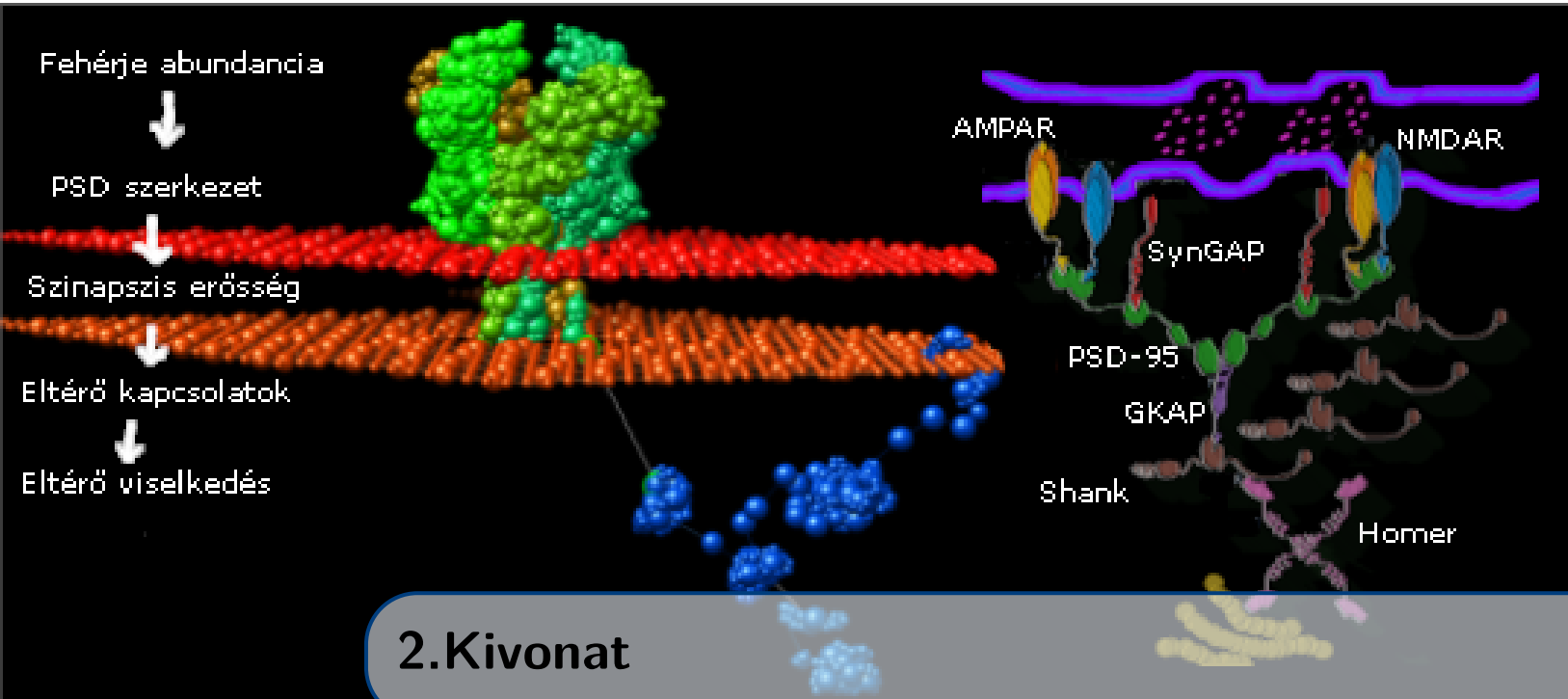


1. Rövidítésjegyzék

rövidítés	angol	magyar/magyarázat
ADHD	Attention deficit hyperactivity disorder	Figyelemzavar
AMPA	-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor	2-amino-3-(5-metil-3-oxo-1,2-oxazol-4-il)propánsav receptor
BLAST	Basic local alignment search tool	Szekvenciainformációt kereső algoritmus
CAD-score	Contact Area Difference	Kontaktfelszín mérőszám
CAMKII	Calmodulin-dependent protein kinase II	Molekula
CBC	Cerebellar cortex	Kisagy kéreg
EVH1	Ena/VASP Homology 1	Homer domén
FOXS	Fast X-Ray Scattering	Szerkezetvizsgáló módszer
GO	Gene Ontology	Génfunkciók tudástára
H4	Homer tetramerization	Homer tetramerizációt tartalmazó szimulációk

rövidítés	angol	magyar/magyarázat
H4SM	Homer tetramerization and shank multimerization	Homer tetramerizációt és shank multimerizációt is tartalmazó szimulációk
ITC	Inferolateral temporal cortex (area TEv, area 20)	Inferolateral temporális kéreg
KO	Knock out	Kiütött - gének kiütésére használt kifejezés
LTP	Long term potentiation	Szinspszis megerősödés aktivitás hatására
M1C	Primary motor cortex	Elsődleges motoros kéreg
MAGUK	Membrane-associated guanylate kinase	Fehérjecsalád
MALDI-TOF	MatrixAssisted Laser Desorption/Ionization Time-of-Flight	Tömegspektrometriás módszer
MFC	Medial prefrontal cortex	Mediális prefrontális kéreg
mRNS		Hírvivő ribonukleinsav
NeEMO	NEtwork Enthalpic MOdelling	Stabilitás predikáló módszer
NMDAR	N-Methyl-D-aspartic acid receptor	N-metil-d-aszparaginsav receptor
NMR	Nuclear magnetic resonance	Mágneses magrezonancia
PALM	Photoactivated Localization Microscopy	Fotoaktivált lokalizációs mikroszkópia
PCA	Principial Component Analyses	Főkomponens Analízis
PDB	Protein data bank	Fehérjeadatbázis

rövidítés	angol	magyar/magyarázat
PDZ	Domain type	Gyakori és jól definiált doméntípus
PSD	Postsynaptic density protein 95	Posztszinaptikus denzitás
PSD-95	Postsynaptic density protein 95	Állványfehérje
PSINDB	Postsynaptic Interaction Database	Interakciós adatbázis a posztszinaptikus denzitás fehérjékre fókuszáltan
RNS-Seq		RNS-szekvenálási módszer
S1C	Primary sensory cortex	Elsődleges szenzoros kéreg
SAM	Sterile alpha motif	Alfamotívumokat tartalmazó Shank domén
SAXS	Small Angle X-ray Scattering	Szerkezetvizsgáló módszer
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis	SDS-polyacrylamide gélelektroforézis
STC	Posterior (caudal) superior temporal cortex (area 22c)	Elülső, felső temporális kéreg
STED	Stimulated Emission Depletion	Stimulált emissziós depléción - szuperrezolúciós képalkotó eljárás
STORM	Stochastic Optical Reconstruction Microscopy	Sztokasztikus optikai rekonstrukciós mikroszkópia
STR	Striatum	Agyi régió
SynGAP	Synaptic Ras GTPase-activating protein	Fehérje
Uniprot		Adatbázis
V1C	Primary visual cortex	Elsődleges vizuális kéreg



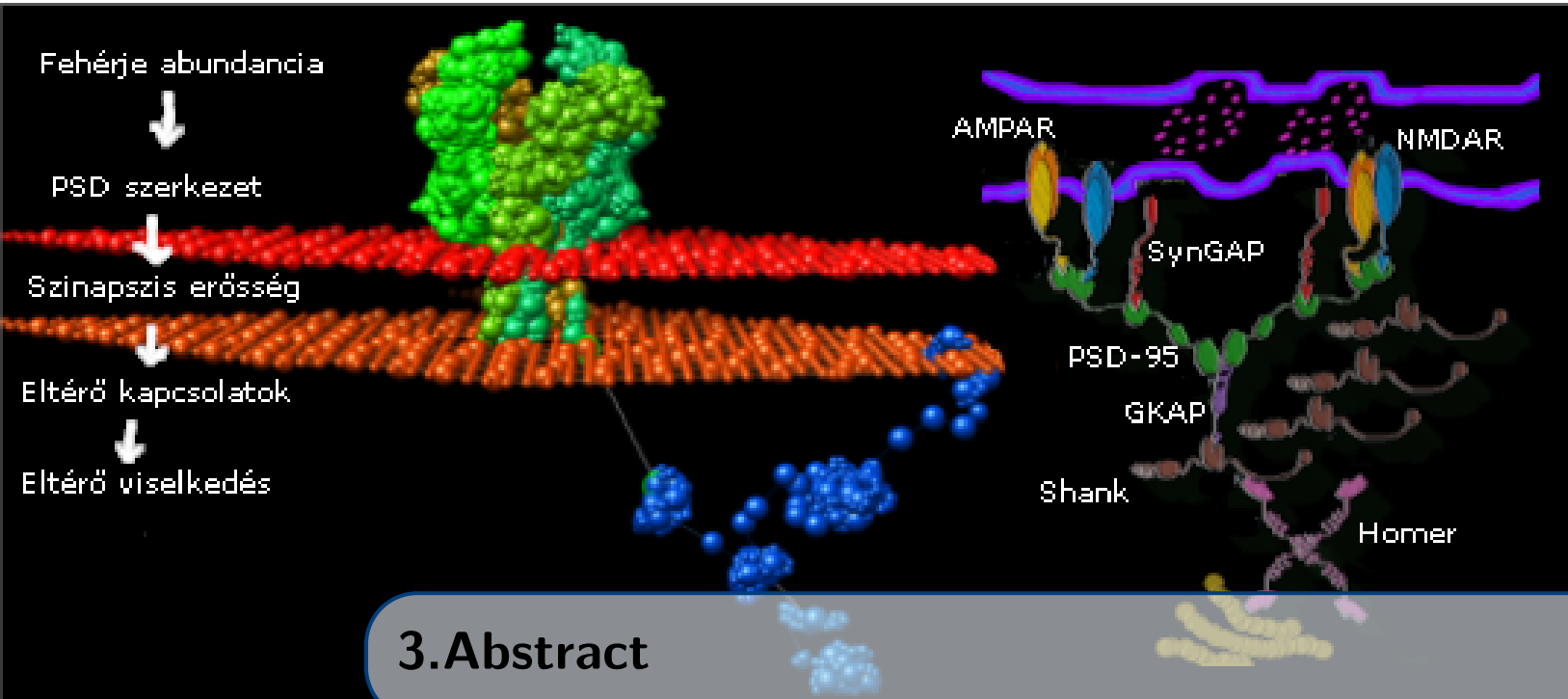
2.Kivonat

A posztszinaptikus denzitás (PSD) egy sűrű fehérjehálózat, amely kulcsszerepet játszik az információfeldolgozásban a tanulás és a memória során, és számos neurológiai rendellenességben is kimutatható a szerepe. A fehérjehálózat bonyolult, a membrán alatt helyezkedik el és ugyanazokból a fő fehérjékből épül fel, de pontos összetételében és felépítésében különbözik a szinapszisok között. Molekuláris szerveződésének részletes jellemzésére irányuló erőfeszítéseket megnehezíti az alkotó fehérjék gyakoriságának nagy változékonysága mind térben, különböző agyterületeken, mind időben, a fejlődés során, a cirkadián ritmusban és a különböző ingerekre adott válaszokban is. Az ebben a hálózatban általánosan előforduló fehérje-fehérje kölcsönhatásokat megzavaró mutációk specifikus hatásokhoz vezethetnek, és a biokémiai hatások rendszerszintre történő transzformációja különösen nagy kihívást jelenthet.

Ebben a tanulmányban a fehérjekötési események nagyszabású sztochasztikus szimulációit futtattam le, hogy előre jelezzem a PSD komplexek jelenlétét és eloszlását. Hét fő PSD fehérje (NMDAR, AMPAR, PSD-95, SynGAP, GKAP, Shank3, Homer1) kölcsönhatásait, beleértve egy vad típusú vagy egy hipomorf PDZ mutáns Shank fehérjét, szimuláltam 22 különböző agyterületen és 42 betegen korábban publikált, kísérletileg meghatározott fehérjegyakorisági adatok alapján (összesen 524 különböző szimuláció).

Eredményeim azt mutatják, hogy a megjelenő fehérjekomplexek relatív aránya érzékeny lehet a fehérjegyakoriság apróbb változásaira is, így az explicit szimulációk felbecsülhetetlenek a fehérje elérhetősége és a komplexképződés közötti összefüggések megértésében. Megfigyeléseink kompatibilisek azzal a forgatókönyvvel, amelyben nagyobb szuperkomplexek jönnek létre a PSD fehérjék rendelkezésre álló kisebb bináris és ternáris asszociációiból. Pontosabban, a Homer1 és Shank3 önasszociációs reakciók lényegesen elősegítik nagyon nagy fehérjekomplexek kialakulását. A leírt szimulációk az első közelítést jelentik a PSD komplex gyakoriságának értékeléséhez, és mint ilyenek, jelentős egyszerűsítéseket alkalmaznak. Ezért közvetlen biológiai relevanciájuk korlátozott lehet, de úgy gondolom, hogy a főbb kvalitatív eredmények hozzájárulhatnak a posztszinapszis molekuláris jellemzőinek megértéséhez. Továbbá eredményeim azt sugallják, hogy az ilyen hipomorf mutációk hatása erősen függ

a teljes hálózat fehérjekomponenseinek elérhetőségétől. Különböző sejttípusok kísérleti adatai alapján szimuláltam a fehérje komplex képződést, eltérő fehérjegyakoriság mellett. A szimulációk kimutatták, hogy az egyetlen fehérje kölcsönhatást megzavaró mutációk csak meghatározott sejttípusokban és specifikus komplexekben vezethetnek fehérjekomplex változáshoz. Bár rendszerem jelenleg még messze van a PSD átfogó leírásától, az eredmények egy olyan módszert javasolnak, ahol egy összetett kereten belül értelmezzük a specifikus mutációkat.



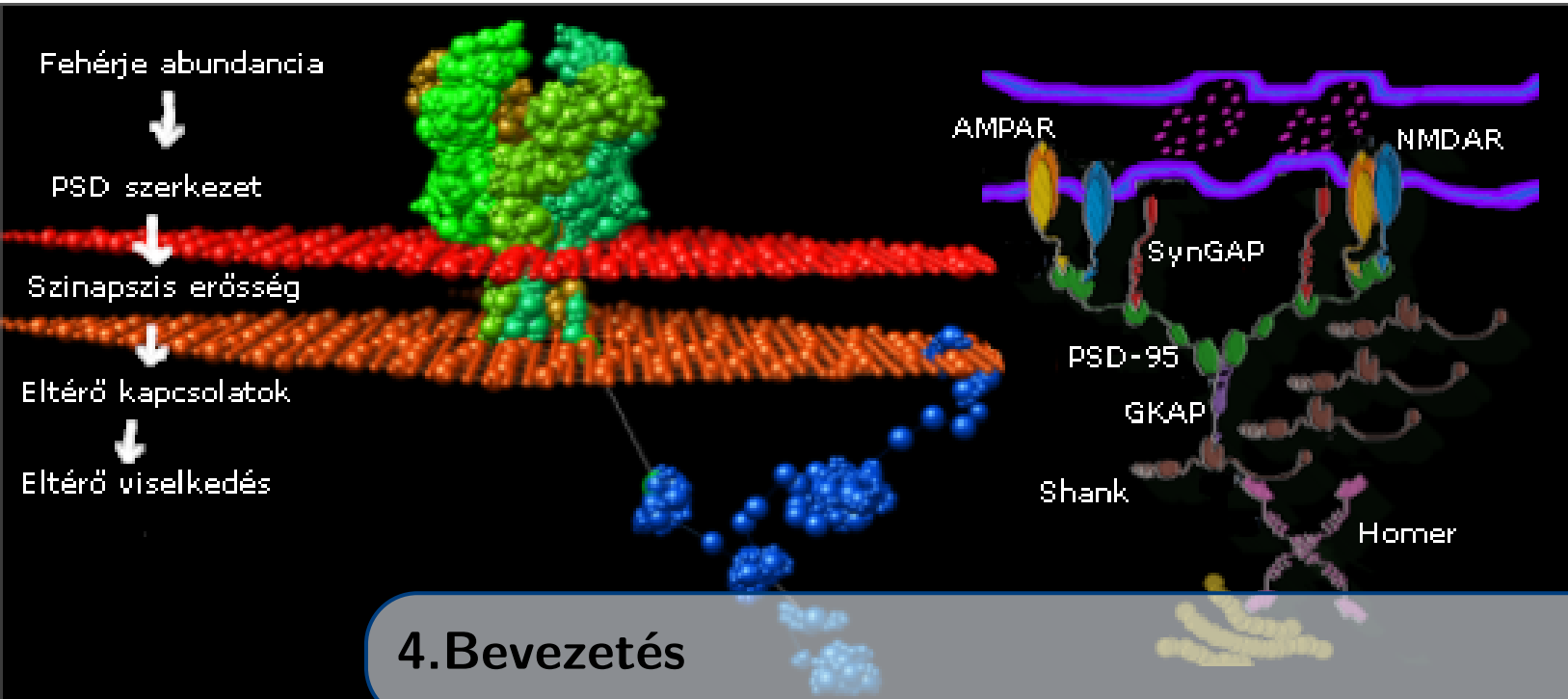
3. Abstract

The postsynaptic density (PSD) is a dense protein network that plays a key role in information processing during learning and memory and is also indicated in many neurological disorders. The protein network is complex, located under the membrane and made up of the same main proteins, but its exact composition and structure differ between synapses. Efforts to characterize its molecular organization in detail are complicated by the great variability of the abundance of the constituent proteins both spatially, in different brain areas, and temporally, during development, in the circadian rhythm and in responses to various stimuli. Mutations that disrupt protein-protein interactions commonly occurring in this network can lead to specific effects, and the transformation of biochemical effects to the systemic level can be particularly challenging.

In this study, I ran large-scale stochastic simulations of protein binding events to predict the presence and distribution of PSD complexes. I simulated the interactions of seven main PSD proteins (NMDAR, AMPAR, PSD-95, SynGAP, GKAP, Shank3, Homer1), including a wild-type and a hypomorphic PDZ mutant Shank protein, based on previously published, experimentally determined protein abundance data in 22 different brain areas and 42 patients (a total of 524 different simulations).

My results show that the relative proportion of emerging protein complexes can be sensitive to small changes in protein abundance, so explicit simulations are invaluable in understanding the relationship between protein availability and complex formation. Our observations are compatible with a scenario in which larger supercomplexes are formed from the available smaller binary and ternary associations of PSD proteins. Specifically, Homer1 and Shank3 self-association reactions significantly promote the formation of very large protein complexes. The described simulations are first approximations to estimate the abundance of the PSD complex and, as such, make significant simplifications. Therefore, their direct biological relevance may be limited, but I believe that the main qualitative results can contribute to the understanding of the molecular features of the postsynapsis. Furthermore, my results suggest that the effect of such hypomorphic mutations strongly depends on the availability of protein components of the entire network. Based on the

experimental data of different cell types, I simulated protein complex formation with different protein abundances. The simulations showed that mutations disrupting a single protein interaction can lead to protein complex changes only in specific cell types and specific complexes. Although my system is currently far from a comprehensive description of PSD, the results suggest a way to interpret specific mutations within a complex framework.



4. Bevezetés

Az emberi viselkedés rendkívül változatos, és egészen a neurotipikus megjelenéstől a neurodivergens autizmus spektrumig terjed, amelyben számos eltérő viselkedésmintát és gondolkodásbeli sajátosságot találunk. Ezek a különbségek gyakran kihívást jelentenek az emberi kapcsolatokban és a társadalmi interakciókban, mivel nincs egyetlen egységes magyarázat ezekre a változásokra. Azonban a neurobiológiai kutatások és a neuroszociális tudományok előrelépései segíthetnek jobban megérteni ezeket a viselkedési különbségeket, és hozzájárulhatnak a megértéshez és az együttműködés javításához azok között, akik eltérő viselkedésmintákkal és gondolkodási sajátosságokkal rendelkeznek. Az oktatás iránti elkötelezettségem is személyesen érdekeltté tesz az agy működésének megismerésében. A neuropedagógia is szeretné megismerni a tanulás és eltérő gondolkodás folyamatait, hogy azt felhasználhassa a tanteremben.

A poszt-szinaptikus denzitás feltérképezése kulcsfontosságú ahhoz, hogy megismerjük azt, hogy az autista viselkedés széles spektruma miképpen alakul ki. Nemcsak a fehérjék mennyisége befolyásolhatja ezt a spektrumot, hanem a fehérjék közötti kölcsönhatások és kötődések is, mivel bizonyos információk a membránreceptoroktól az idegsejt belső területeire az úthálózatként szerteágazó összekapcsolt fehérjekomplexeken keresztül juthatnak.

A kutatómunkám tartalmazta az irodalomkutatást, a Cytocast program használatát, megismerését és tesztelését, javaslatok írását a fejlesztőknek (Keömley-Horváth Bence Márknak), nagy mennyiségű adatok kezelését, az adatok kezeléséhez többnyire Python kódok írását, szimulációk bemeneti paramétereinek megtervezését az irodalomban fellelhető kísérletes adatok figyelembevételével, szimulációk lefuttatását, valamint a kapott eredmények kiértékelési metodikájának kidolgozását és az értékelés elvégzését.

4.1 A poszt-szinaptikus denzitás

A szinapszisok az idegsejtek közötti kapcsolódási pontok, ahol az információ továbbítódik egy neuronról a másikra. A kémiai szinapszisoknak három fő része van: preszinaptikus neuron, amely az információ küldő része [1], a gliasejtek, amelyek az információ szabályozó

része [2] és a minket érdeklő posztszinaptikus idegsejt, amely az információ fogadó része.

A posztszinaptikus denzitás (postsynaptic density, PSD) a posztszinaptikus neuron dendritjeinek membránjához közeli fehérjedús része [3]. Elektronmikroszkópos képeken sötét („elektronenz”) sávként jelenik meg, nevét is erről kapta. A PSD fehérjéi között számos receptor és más állványfehérje található, amelyek kulcsfontosságúak az elektrokémiai jelátvitel folyamatában [4]. Ennek fontosságát az a korreláció is jelzi, hogy hosszútávú aktivitás esetén (long term potenciáció (LTP)) a posztszinaptikus denzitás mérete megnövekedik, erősebbé válik [5]. A fehérjehálózat reaktív átrendeződését a tanulás és memória fontos folyamatának tekintjük [6].

A posztszinaptikus denzitást számos módon vizsgálták. Tartalmának meghatározására többek között SDS-polyacrylamide gélelektroforézist (SDS-PAGE) használtak kiválasztva fehérjéket, MALDI-TOF tömegspektrométerrel egybekötve [7]. Ezekkel az eljárásokkal meg tudják találni az építőelemeit a hálózatnak, azonban a hálózat tényleges felépítését, klaszterezettségét és dinamikus reaktív átszerveződését nem tudják vizsgálni. Ettől függetlenül elengedhetetlen része volt a régió feltárásának. Munkám is kiindulási alapnak tekinti, hogy „ismerjük az építőköveket és azok mennyiségét”.

A PSD-ben (posztszinaptikus denzitásban) lejátszódó számos folyamatról rendelkezünk már valamennyi ismerettel [8], [9]. Patterson et al. (2011) kutatásai kimutatták, hogy a PSD-ben zajló fehérje-protein interakciók és foszforilációs események kulcsfontosságúak a szinaptikus plaszticitás és jelátvitel szabályozásában. Soria-Fregozo et al. (2012) pedig bemutatták, hogy a PSD-ben lévő strukturális fehérjék, mint például a PSD-95, alapvető szerepet játszanak a szinapszis stabilitásában és az ionszűrő eloszlásában. Közéjük tartozik a NOS fehérje által elindított apoptózis [10], valamint a kalcium-ion reaktív felszabadulása is [11]. A PSD mint teljes hálózat szerkezetéről számos erőfeszítés ellenére csak viszonylag korlátozott ismereteink vannak. Egyes eredmények arra utalnak, hogy szerepet játszik benne az ún. fehérje fázisszeparáció jelensége. A fázisszeparáció gyakran fordul elő, amikor egy keverékben lévő komponenseknek különböző kémiai vagy fizikai tulajdonságai vannak [12], és ezáltal a keverék nem marad homogén [13]. A fázisszeparáció jelentősége, hogy az adott receptorok is ezek alapján klasztereződnek a membrán felszínén [14]. A receptorok eltérő térbeli eloszlása eltérő valószínűségi mezőket eredményez, amely miatt a vevő neuron reakciója is megváltozik "ugyanazon" jel hatására. Hagyományosan úgy vélik, hogy elsősorban az aktivitás okozta posztszinaptikus AMPA (α -amino-3-hidroxi-5-metil-4-isoxazolepropionikus acid) receptorok számának növekedésén keresztül valósul meg a hosszú távú potenciáció (LTP), a szinapszisok jelentős erősödését tapasztalhatjuk a szinaptikus összetevők térbeli eloszlásának átrendezésével, anélkül, hogy növekedne a receptorok száma [15]. Ez a fajta receptoreloszlás mediálható és megváltoztatható, például CAMKII molekula mediátorként viselkedik, melyet szuper-rezolúciós mikroszkóppal vizsgáltak [16]. Elmondható, hogy a szuperrezolúciós képalkotó eljárások már a fehérjék (és ionok) pozíciójáról is adnak információkat közelebb vive a tényleges szerkezethez [17]. A receptorok nanodoménjeibe történő szerveződését megfigyelhetjük. Az AMPAR lokalizációjának nagy sűrűségű és szuperfelbontású képével megmutatták azt, hogy az AMPAR nanodoménekbe szerveződik [14].

Tekintve, hogy a posztszinaptikus denzitás egy hatalmas, nagyon sok fehérjét magába foglaló komplex rendszer, ezért az irodalomkutatás részeként ki kellett választanom a posztszinaptikus denzitásról ismert komplexrendszerek közül egy kutatandó részegységet, amellyel elkezdhettem az eljárási lánc felépítését. A választásom a Feng és kutatótársai

által vizsgált fehérjehálózatára esett [18], mert az irodalom ezt a hálózatot jól leírja és mélységében a receptoroktól a citoszkeletonig eljutunk. A hálózat hét alapvető fehérjéből állt, két különböző típusú receptorból, melyek a PSD-95 tandem-PDZ doménjéhez kötődnek. A PSD-95-höz továbbá egy további kötőhelyen a SynGAP kötődik, mely befolyásolhatja, mely más fehérjék férnek hozzá a maradék szabad doménjéhez a fehérjének. Ilyen lehetséges fehérje, ami a PSD-95-höz tud még kötni, a GKAP fehérje, amihez köt a későbbiekben lényeges Shank3 fehérje, majd ahhoz a Homer1 fehérje. A Homer1 fehérje továbbá az aktinfilamentumokhoz is tud csatlakozni, így létrehozva a neuron belsejével a teljes összeköttetés lehetőségét.

Nyilvánvalóan még ezek sokfélesége sem ragadható meg a szimulációimban 7.3, mivel az összes fehérjének több izoformája van, eltérő partnerkötő tulajdonságokkal, míg én csak az UniProt adatbázisban szereplő reprezentatív izoformákat vettem figyelembe. Ráadásul mindezeknek a fehérjéknek további kötőpartnerei vannak, amelyek közül feltehetően jó néhányat nem is ismerünk még. Ennek ellenére úgy gondolom, hogy ez a részalmaz megfelelő jelölt az első szimulációs tanulmányhoz az itt leírtak szerint.

4.2 A szinaptom elmélet

A szinaptom elméletet Seth Grant és kutatócsoportja dolgozta ki [19]. Ez az elmélet azt az elgondolást módosítja, bővíti ki, miszerint egyelőre sok biológia tankönyv úgy tartja, hogy a különböző agyi funkciókat a neuronok különböző vezetékéke okozza. Ennek ellenére a vezetékek, neuron-neuron kapcsolatok nem adnak teljes választ a genetikai eredetű neurodegeneratív rendellenességekre, mint például az Alzheimer és az Autizmus spektrum.

A centrális dogma szerint a gének nem közvetlenül kódolják a fehérjéket, hanem először RNS-re íródnak át, majd az RNS fordítódik le fehérjévé [20]. A gének tehát nem közvetlenül határozzák meg a neuronok nyúlványait (azaz az axonokat és dendriteket), hanem a kódolt fehérjék révén befolyásolják azok kialakulását és funkcióját. A szinaptom elmélet szerint a különböző neurontípusokat az expresszálandó fehérjék specifikus mintázata határozza meg, mivel ezek a fehérjék végső soron meghatározzák a sejtek szerkezetét és funkcióit [19].

Tehát a szinaptom elmélet szerint az adott neurontípust az expresszálandó fehérjék határozzák meg. Így a fehérjék közvetve befolyásolják az agyi funkciókat és képesek a viselkedést megváltoztatni. Ennek az elméletnek a bővebb megismerését és vizsgálatát tűztem ki PhD Disszertációm céljává számítógépes szimulációk és modellek segítségével, ugyanis az elmélet érdekessége, hogy kitér magára a tanulás folyamatára is [21]: miszerint a tanulás a szinapszis proteomjának módosításával történik [22], ez nem igényli a szinaptikus súly hosszú távú erősödését (LTP) vagy az új szinapszisok növekedését, és az elmélet azt jósolja, hogy az LTP modulálja az információ visszahívását [21]. A későbbi háromdimenziós modellezések szempontjából az érdekes, ahogy a Seth Grant is megjegyzi [19]: a szinaptom, fehérjehálózatok térbeli architektúrája abból a mögöttes molekuláris hierarchiából származik, amely a genom által kódolt fehérjék szupramolekuláris összeállításával komplexekké és szuperkomplexekké kapcsolja. Ez a molekuláris hierarchia megmagyarázza, hogy a genom szintjén zajló evolúció hogyan eredményezi a szervezet viselkedési repertoárjának megváltozását. Az ezt a molekuláris hierarchiát megszakító mutációk megváltoztatják a szinaptom fehérjehálózatok felépítését, potenciálisan figyelembe véve a neurológiai és pszichiátriai rendellenességekkel kapcsolatos viselkedési fenotípusokat [23], [24]. Tehát a kapott komplexek térbeli elrendeződésének vizsgálata lesz a végső célom.

A bemutatott szimulációkat nem csak agyi régió típusokra lehet elvégezni, hanem ugyanúgy el lehetne végezni neuron típusok megvizsgálására. A szakirodalomban már megjelentek ilyen vizsgálatok, melyekben a neurontípusok családjait próbálják meg a transzkriptom alapján beazonosítani fák generálása segítségével [25], [26]. Azt már belátták, hogy transzkriptom alapján a nagyobb családtípusokon belül folytonos variancia figyelhető meg [27]. Ez a megfigyelés a szinaptom elméletet támasztja alá és azt, hogy a fehérjekoncentráció (és nanodomén szerveződésük) valóban finomhangolja a neuronokat, ezáltal a neuronok nagy diverzitását okozva, akár neuroncsaládon belül is [28], [29]. Ez a hatás a posztzinaptikus denzitáson keresztül érvényesülhet.

4.2.1 Fehérjegyakoriságok a posztzinaptikus denzitásban

Az általam elvégzett szimulációk során kritikus bemeneti adatként szerepel az egyes fehérjék konkrét darabszáma, gyakorisága. Az irodalomban kevés kivételtől eltekintve RNS-expressziós adatokat lehet találni erre vonatkozóan [30]. Természetesen kapcsolat van az RNS expresszió és a fehérje mennyisége között, viszont ez a kapcsolat nem feltétlenül egy az egyben lineáris [31] és nagyon sok behatástól függ: az RNS transzportjától [32], [33] a translációján [34] és stabilitásán [35] át egészen a fehérje lebomlásáig [36] lásd a 4.1. ábrán. Amennyiben közvetlenül a fehérjék gyakoriságát használjuk, akkor természetesen az előző különböző folyamatokból (transzláció, folding és transzport stb.) keletkező hibákat könnyebben ki lehet szűrni viszont másfajta kísérletes eljárás kell az adatok megszerzéséhez.

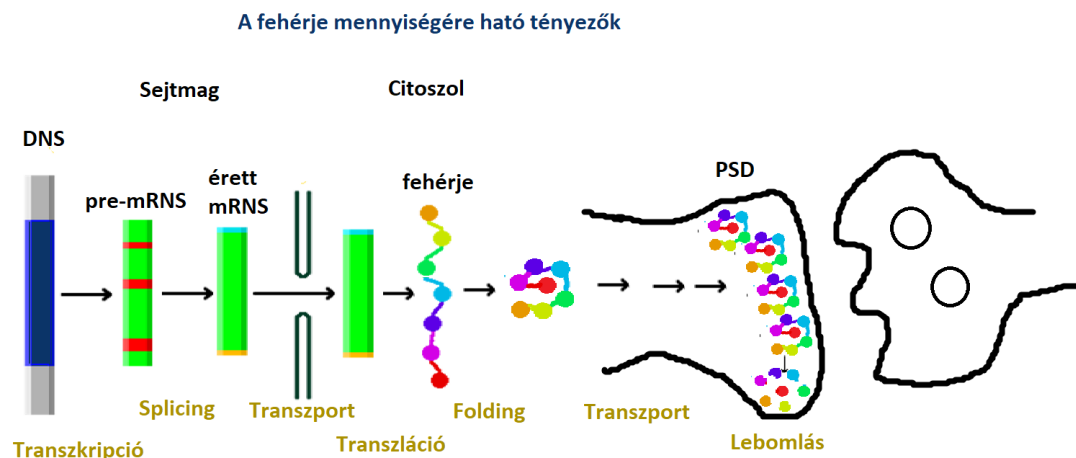
Az expressziós szint és a fehérjegyakoriság közötti összefüggést úgy lehetne vizsgálni, hogyha ugyanazon sejtben egyszerre ki tudnánk nyerni az expressziós adatokat és a tényleges fehérjegyakorisági adatokat is, majd ebből egy paraméteres függvényt fel tudnánk írni és a paramétereket pl. maximum likelihood becsléssel meg tudnánk becsülni. Mindazonáltal az egyes fehérjék egymás expressziós hatékonyságát is befolyásolják, a nagymennyiségű adatokat figyelembe véve, gépi tanulásra lehet szükség a paraméterek megbecsléséhez.

4.3 A posztzinaptikus denzitást vizsgálatára használt kísérletes módszerek

4.3.1 Fehérjeösszetétel vizsgálatára alkalmas módszerek

Az SDS-PAGE, Western blot és MALDI-TOF technikákat integráltan alkalmazza kutató csoportunk a kutatások során a fehérjék azonosítására, kvantifikálására és funkcionális elemzésére.

Az SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) segítségével a különböző (5 és 250 kDA között) méretű és töltésű fehérjéket el lehet különíteni egymástól [37]. A posztzinaptikus denzitás elemzésének kiinduló kérdése, hogy milyen fehérjék találhatóak meg benne, Western Blottal együtt specifikus fehérjék meglétét igazolhatjuk azáltal, hogy a méret szerint elkülönített fehérjéket egyben specifikusan meg is jelöltük - a megfelelő méretnél megfelelő jelölőt kell látnunk. Az SDS-PAGE technológia alapelve, hogy az SDS erősen kötődik a fehérjékhez. Amikor az SDS-sel forraljuk a fehérjéket, azok negatív töltést szereznek a molekulatömegük arányában, így vándorolnak az akrilamid gélben a molekulatömegük szerint. Minél kisebb a futó fehérje mérete, annál gyorsabban halad a gélpórusokon keresztül [38]. Ebből adódóan miután megállítjuk a folyamatot, a nagyobb fehérjék magasabban maradnak a kisebb fehérjékhez képest, egy létrát kialakítva molekulatömeg szerint. A folyamatot az 4.2 ábra szemlélteti.



4.1. ábra. **A fehérje mennyiségét befolyásoló tényezők a Centrális Dogmától a degradációig.** Az adott fehérje mennyiségét és tulajdonságait számos tényező befolyásolja. Az ábra a fehérje keletkezését és annak útját szemlélteti, amely minden állomásán modulálódhat a fehérje mennyisége a sejtmagtól a poszt-szinaptikus denzitásig.

Az SDS-PAGE által elszapérált fehérjék tovább vizsgálhatóak. A MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight) egy speciális tömegspektrometriai eljárás, amelynek lényege, hogy a vizsgált mintát egy mátrixanyaggal keverve lézerrel gerjesztik. Ennek eredményeként az ionizált minták atomjai leválnak elég energia behatására (deszorpció), és meghatározható a tömegük a repülési idő segítségével, amíg a detektorba eljutnak. [39]

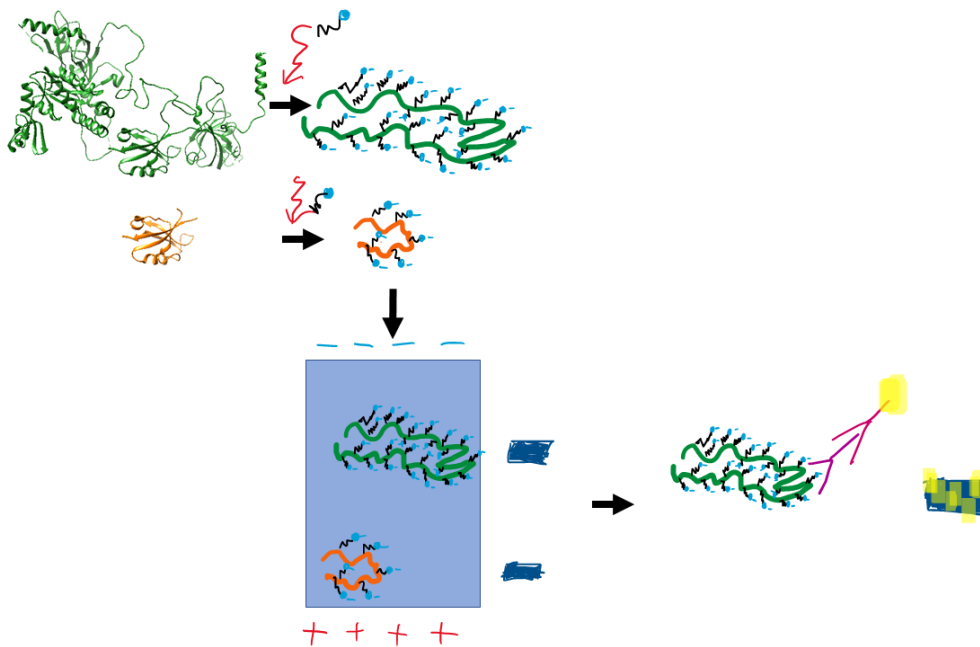
A MALDI-TOF rendkívül hasznos a fehérjék azonosításában és karakterizálásában, mivel lehetővé teszi a fehérjék molekulatömegének gyors és pontos meghatározását. Emellett a MALDI-TOF kombinálható a foszforiláció vizsgálatával is, mert a foszforiláció megváltoztatja a fehérje relatív ionizációját és tömegét [40]. A foszforiláció fontos jelenség a poszt-szinaptikus denzitás esetében is, mivel sok fehérje aktiválódik vagy inaktiválódik ennek eredményeként. A MALDI-TOF segítségével az ilyen és egyéb poszttranszlációs módosulásokat (pl. metiláció) is nyomon lehet követni és azonosítani.

4.3.2 Szuperrezolúciós mikroszkópiai eljárások

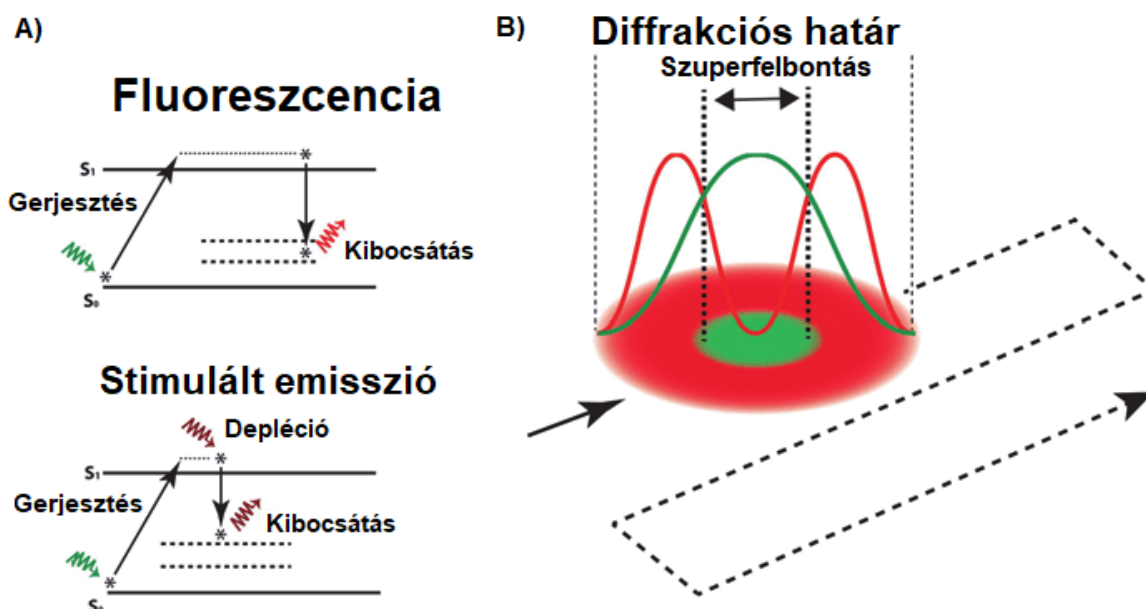
A szuperrezolúciós mikroszkópia egy olyan speciális eljárás, amely lehetővé teszi, hogy a diffrakciós limitnél nagyobb felbontást érjünk el fluoreszcens mikroszkópiával. A STED (Stimulated Emission Depletion) [41] eljárás egy ilyen szuperrezolúciós technika, amellyel olyan olyan morfológiai részleteket tárhatunk fel, amelyek hagyományos fénymikroszkópia esetén nem lennének láthatók.

A STED szuperrezolúciós mikroszkópiál lehetővé teszi a PSD és a kapcsolódó struktúrák részletes vizsgálatát. Ez a technika - noha nem használtam - nemcsak az előzetes információgyűjtésben volt hasznos, hanem az eredmények utóvalidálását is, biztosítja, hogy a szimulációk és a valós biológiai megfigyelések összhangban legyenek egymással.

Alapelve, hogy a minta megvilágításához két fényforrást alkalmaznak (lásd 4.3. ábra):



4.2. ábra. **SDS-PAGE és Western blot folyamata** A zöld egy nagyobb méretű fehérjét, míg a narancssárga egy kisebb méretű fehérjét reprezentál. A negatív töltéssel rendelkező köpeny töltésszáma egyenesen arányos a fehérje méretével. Ezért a nagyobb fehérjék lassabban mozognak a két pólus között a gélben. Létrát képezve. Az adott fehérjét jelölőmolekulákkal megjelölhetjük, hogy fehérjespecifikus jeleket láthassunk és ezáltal megbecsülhessük az azonos méretű fehérjéket tartalmazó csomagban mennyire tisztán található meg a keresett fehérjénk.



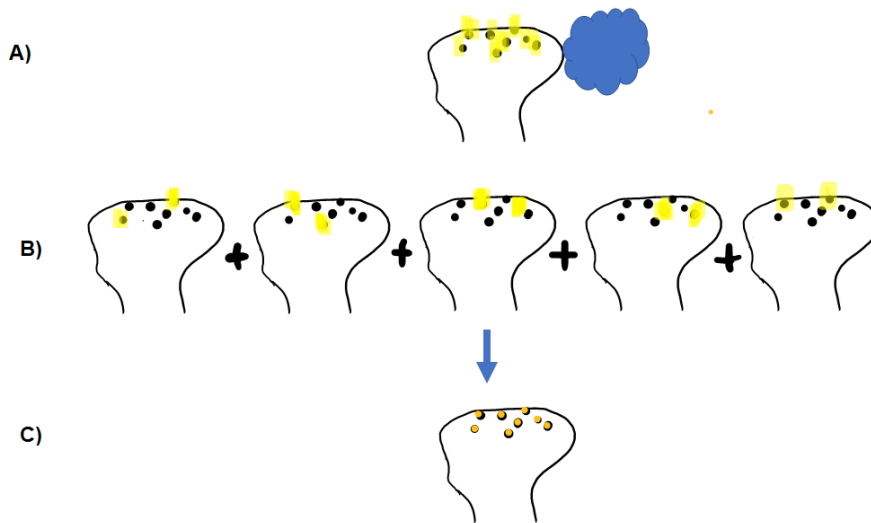
4.3. ábra. Vöröseltolódás a stimulált emisszió hatására

Az ábra az alábbi cikkből származik [43]. a) Stimulált emisszió hatására a foton hamarabb kilép, így kisebb energiája lesz. Vöröseltolódik. b) A vöröseltolódás, a fluoreszcens fókusz haranggörbéi a síkon szemléltetve.

egy gerjesztő lézert és egy vörös eltolódott depléciós lézert, amit STED nyalábnak nevezünk. A fókuszpontban a gerjesztő lézert használjuk - spontán emisszió történik, míg a környező területeken a depléciós lézer még a teljes relaxáció előtt perturbálja a rendszert, magasabb energiaállapotban kényszeríti relaxálni azt, a foton így kisebb energiával kényszerül kilépésre a lézer hatására (stimulált emisszió). Ezáltal a környező, nem fókuszban lévő területekről vöröseltolódott fény szabadul ki, ami így nem befolyásolja a fókuszban lévő rendszeren fluoreszkáló területről alkotott képet [42].

A teljes kép elkészítéséhez a STED lézert előállított középső nulla pontot végigpásztazzák a mintán. A STED módszer fejlesztéséért 2014-ben Stefan W. Hell kapta a Kémiai Nobel-díjat (lásd nobelprize.org).

Fontos megjegyezni, hogy az erős lézertény, ami a STED eljárás során használatos, potenciálisan károsíthatja a mintát, és gyors kifakuláshoz vezethet. Azonban a szuperrezolúciós mikroszkópia, és a STED technika különösen, nagy segítséget nyújt a biológiai minták részletes vizsgálatában és a mikroszkópia területén előrelépésekhez vezethet. Például megfigyelhető, hogy a SynGAP gén kiiktatásának milyen hatása van a CA1 piramissejtek apikális dendritikus szakaszainak alakjára, amit felületi renderelt STED-képekkel vizsgáltak [44]. Ezen keresztül közvetve megtudhatjuk, hogy milyen hatása van a posztzinaptikus denzitásnak a dendritikus alakváltozásokra. Ugyanígy a nanodomének megfigyelésére is alkalmas [45].



4.4. ábra. **STORM alapelve.** A) A legfelső ábra egy hagyományos fluoreszkáló képalkotást szimbolizál, melyben minden jelölt molekula egyszerre fénylik fel, ezáltal egy zajos fényfelhőt létrehozva, mert a közeli fényforrások nem válnak elkülöníthetővé. B) A második sor bemutatja, hogy STORM eljárás során véletlenszerűen aktíváljuk a fényforrásokat és több képet készítünk, ezáltal a túl közel eső fényforrások egyszerre nem aktiválódnak - nem zavarják meg egymás jelét. Ezáltal a képek külön-külön megvilágítják a kívánt molekula pontos pozícióját. C) Végül a STORM ábra ezen képek összege lesz (utolsó sor), ahol minden molekula külön látható, még a közeli is.

4.3.3 PALM és STORM

A PALM (Photoactivated Localization Microscopy) és a STORM (Stochastic Optical Reconstruction Microscopy) módszer az egyes, fotoaktiválható tulajdonságokkal rendelkező fluoroforok sztochasztikus aktiválásán alapul. Lehetővé teszi egyéni molekulák térbeli elhelyezkedésének magas pontosságú térbeli felbontású megjelenítését még sűrű populációkban is. Ezzel ideális módszert teremtve arra, hogy a poszt-szinaptikus denzitásról fehérjeszintű felbontású képeket készíthessünk. Ezzel a felbontással egymáshoz közel eső fehérjéket is külön láthatunk és ezekről feltételezhetjük, hogy kapcsolatban állnak egymással. Például STORM segítségével egyazon képen láthatóvá tették a Bassoon és a poszt-szinaptikus Homer fehérjéket eltérő jelölést használva [46].

Az eljárások lényege, hogy nagyszámú képet készítünk, és ennek során randomizáljuk, hogy éppen mely jelölők aktiválódnak, így várhatóan az egymás melletti jelölők egy adott képen nem aktiválódnak egyszerre, így nem mossák össze fényüket. A több különböző képet összeadva megkaphatjuk a magasfelbontású képet, melyen minden egyes jelölő elkülönülten látszik annak ellenére, hogy eredetileg annyira közel voltak egymáshoz, hogy hagyományos képalkotó eljárással már a diffrakciós limit miatt nem lehetett volna külön észlelni őket - lásd 4.4.

4.3.4 Génkiütéses viselkedésvizsgálatok

A knock-out (KO) egerekben génkiütés során viselkedésvizsgáló tesztek segítségével a fehérjék hiányával összefüggő viselkedési változásokat vizsgálják. Ez hasznos iránytű lehet a PSD kutatásában, de nem feltétlenül nyújt egyértelmű bizonyítékot a mechanizmusokra vagy a PSD működésének részleteire nézve.

A KO kísérletekből származó irodalom kritikus szerepet játszik abban, hogy meghatározzuk, mely fehérjék mutációit érdemes részletesen vizsgálni a kutatás során. Ezek az információk segítenek azonosítani a funkcionálisan fontos fehérjéket, megérteni a fehérje-kölcsönhatásokat

Kimutatták, hogy a Shank3-hiányos egerek bizonyos mértékű hiperaktivitást mutatnak [47], amely azért lehet érdekes, mert ez a jellemző az autizmus bizonyos eseteiben jelen van. Az elvégzett tesztek közé tartozik a Zero Maze teszt és a Beam Walking teszt, amelyek a viselkedés és motoros készségek értékelésére szolgálnak.

A Zero Maze teszt az állatok szorongását és félelmüket méri [48]. A labirintus négy szakaszból áll, amelyek egy kör alakú pályán helyezkednek el. A pálya körülbelül egy lábbal a levegőben függ. Két szakaszon magas, keskeny falak találhatók, amelyek általában sötét színűek. A patkányok általában ezeket a szakaszokat részesítik előnyben, mert úgy érzik, elrejtőzhetnek a látótérből. A másik két szakasz pedig nyitott, fal nélküli terület, amelyeken nincsen lehetőség elrejtőzni. Mivel a patkányok és egerek, akik nyitott területekre kerülnek, gyakran a prédaállatok számára könnyű célponttá válnak, ösztönösen kerülnek ezeket a területeket. Az állatok szorongásának szintjét az alapján mérhetik, mennyi időt töltenek el mindegyik szakaszon. Ezenkívül megfigyelhetőek az állatok mozgási mintái és sebességük is, amelyet az említett cikkben is tettek.

Ezzel együtt a Beam Walking teszt a mozgás- és egyensúlyzavarokat vizsgálja [49]. Az állatokat először egy keskeny deszka egyik sarkában helyezik el, majd legalább három alkalommal átengedik őket ezen a keskeny deszkán a másik végéig sétálni. A deszka 1-3 cm széles és az állatok otthoni ketrecéhez vezet, így vonzza őket a célpont felé. Minden próbában rögzítik, hány félre lépés történt, és mennyi idő alatt sétáltak át a deszkán. Ezzel a módszerrel nemcsak az egyensúlyzavart, hanem a sebességet, hezitációt és mozgási kényszereket is figyelemmel kísérhetjük.

Szeretném ismét kiemelni, hogy bár nem vagyunk képesek közvetlenül megfigyelni azt a pontos mechanizmust, amely mindezt előidézi, ezek a kísérletek fontos alapját képezik az én kutatásaimnak is. Segítenek előre jelezni, hogy mely fehérjék mutációja vagy gyakoriságváltozása lehet érdekes és érdemes közelebbről vizsgálni a szimulációimban. Reméljük, hogy ezáltal közelebb juthatunk a hatásmechanizmus feltárásához is. Azt is észben kell tartani, hogy pontosan nem tudjuk, milyen hatások vezetnek adott kimenetelhez - pl. nem biztos, hogy az LTP-t vizsgáljuk általa.

4.3.5 Fehérjék térszerkezetének meghatározása

A posztszinaptikus denzitás szerveződésének igazán mély megértéséhez tudnunk kell, hogy az adott építőelemek, fehérjék térbeli entitásokként hogyan jelennek meg, Ez több okból sem triviális probléma: a fehérjék belső dinamikával rendelkeznek [50], így valójában térszerkezetük folyamatosan változik és alkalmazkodik is. Fehérjék belső dinamikájával foglalkoztam BSc-m során a GROMACS segítségével. A másik ok egyben az, hogy a fehérjék "túl nagy" molekulák, emiatt kísérletes technikákkal szerkezetük egésze nehezen vizsgálható

(habár törekvések ezen határ átlépésére vannak) [51]. Ez különösen igaz a PSD nagyméretű állványfehérjéire, de a több alegységből felépülő transzmembrán receptorokra is. Emiatt legtöbbször NMR spektroszkópiával vagy röntgenkrisztallográfiával ezen fehérjék csak egy-egy jól definiált szakaszát, doménjét tudjuk megvizsgálni - funkcionalitás és szerkezeti egység szintjén így tipikusan csak a globuláris domének szerkezeteiről van információnk. Teljes fehérjeszerkezet becslésére, predikciójára többféle módszer is rendelkezésünkre áll, de jelenleg a legfrissebb áttörést az AlphaFold2 nevű neurális hálózatban alkalmazott mélytanulási algoritmus hozta. Az algoritmus lehetővé teszi bármely adott fehérje háromdimenziós szerkezetének predikcióját csupán az aminosavak sorrendjének ismeretében [52]. Ennek eredményeként szinte bármely fehérjének készíthetünk megbízható számítógépes szerkezeti modellt. A modell nem pontos [53], de tovább pontosítható a kísérletes adatok ismeretében molekulapótlási modell segítségével [54], amennyiben találunk a fehérjéhez hasonló ismert térszerkezetet és azt templátként ráillesztjük az eredményekre. "Az AlphaFold előrejelzései ugyan rendkívül pontosak lehetnek, de még a magabiztosnak tűnő részek is ellentmondhatnak a kísérleti adatoknak, így azokat hipotézisként kell kezelni." [55]. Külön kiemelendők az oligomerizációs motívumok (és kaméleon szekvenciák [56]), pl. coiled coil szakaszok, melyeket az AlphaFold2 természeténél fogva, mind monomerekre kitalált módszer, nem képes kezelni, ennek eredményeképpen pl. a Homer fehérjék esetében a teljes szerkezetre tett becslése nem használható [57]. Ezen túlmenően az egyes globuláris domének relatív orientációja is pontatlan lehet, és ez is kritikus szempont a posztszinaptikus denzitás komplexek kialakulásának és szerkezetének vizsgálatakor. Többek között azért, mert az orientáció megváltozhat kötés hatására [58], [59]. Ezért ezeket a modelleket kellő óvatossággal kell kezelnünk, és figyelembe kell vennünk a domének orientációjának potenciális hibáit is a kutatás során.

4.4 Fehérjekomplexek kialakulásának ágens alapú szimulációja

Számos modellező és szimulációs eljárás van, amellyel a neuronban lévő folyamatok megismerhetők, a jelátvivő folyamatok hálózatának logikai leírásától a neuron aktivációs mintáinak modellezéséig [60]. Kutatásom a fehérjekomplexek szintjén vizsgálja a változásokat.

Fehérjekomplexek keletkezését le lehet írni ahhoz hasonlóan, ahogyan egy kémiai reakciót is: két molekula véletlenszerűen mozog, majd találkozik. Ennek során, ha megfelelőek a körülmények, akkor reagálnak egymással és létrejön az új molekula. Az ilyen jellegű eseményeket a módszerekben ismertetett Gillespie-algoritmussal lehet szimulálni.

A szimulációs módszer lényege, hogy a molekulákat ágens alapon mozgatjuk, majd ha elég közel kerülnek, akkor azt találkozási eseménynek tekintjük. A találkozás során adott valószínűséggel létrehozunk közöttük a kapcsolatot (a két különálló fehérjéből egy összeállt komplex lesz - két ágens átalakul egy ágenssé). Ezt az adott valószínűséget biológiailag a kötési erősségek, affinitások írják le. A komplex ugyanakkor a disszociációs állandó által meghatározott valószínűséggel egy következő lépésben szét is eshet.

Két szoftver alkalmas a fehérjék kölcsönhatásainak ilyen szimulálására, a SiComPre és annak továbbfejlesztése a Cytocast, melyeket alább ismeretetek részletesebben. A Cytocast által használt alapelv a Covid-19-hez kapcsolódó szimulációk alapjául is szolgált [61], ahol az ágensek a személyek, a kötődések a fertőzés átterjedése volt. A módszer költséghatékony és relatíve gyors.

4.5 Fehérjekomplexek 3D szerkezetének modellezése

Az ágens alapú modellezés legnagyobb hátránya, hogy nagy memóriaigénye lehet, így nem reális, hogy a fehérjék tényleges háromdimenziós szerkezetét ill. az ebből adódó megkötéseket egy ilyen szimulációban reprezentálni tudjuk. Egy fehérje, egy komplex ágens egy adott pontként jelenik meg - mégha kötőhelyei virtuálisan léteztek is. Ugyanakkor fehérjekomplexek szerkezetének, alakjának modellezésére is léteznek használható szoftverek. Munkám során fontos szempontnak tartom, hogy a mennyiségi ágens alapú szimulációk eredményeit és a minőségi 3D-s modellezés eredményeit közösen szemlélhessem és lehetővé tegyem a későbbiekben ezek közös folyamatban történő integrálását.

Az IMP (Integrative Modeling Platform) programot nevéből is adódóan az integratív modellezés céljából alkották meg, ami azt jelenti, hogy többféle forrásból származó bemeneti adatot, illetve egy adott rendszer egyes részeinek különböző típusú reprezentációit képes integrált módon figyelembe venni a modellezési folyamat során. A platform lehetővé teszi nagyméretű komplexek háromdimenziós szerkezeti modelljének létrehozását számos adat együttes felhasználásával.

Az IMP Előnyei:

- A domének háromdimenziós orientációja figyelembe vehető.
- A komplexet háromdimenziós formában kapjuk meg.
- Kvalitatív modellt kapunk.
- Nyílt forráskódú.

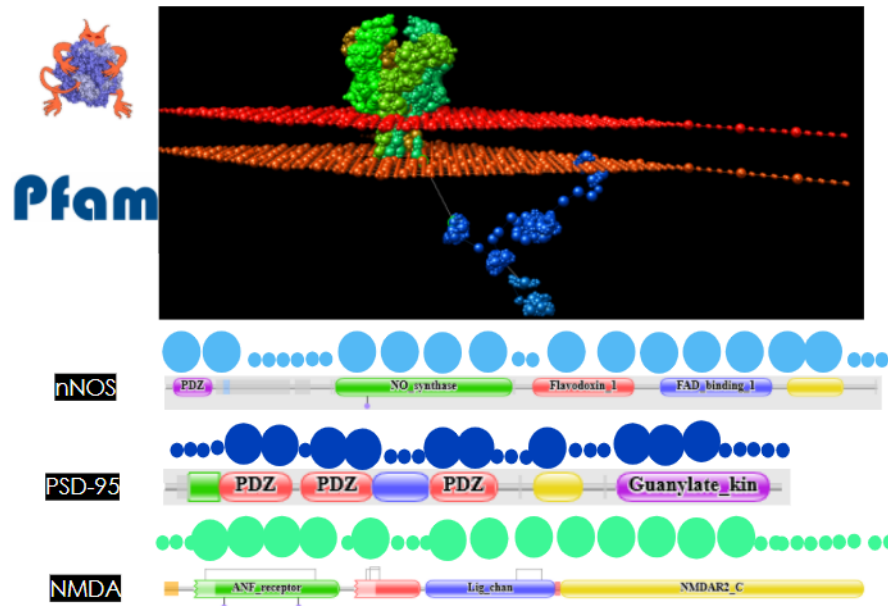
Hátrányai:

- A fehérjék közötti kapcsolatokat sokkal korlátozottabb körben képes figyelembe venni mint az ágens-alapú modellek
- Használata programozási tudást (Python) igényel

Az IMP modellezési folyamatnak öt lépése van [62]:

- Az első lépés a saját eljárási folyamatomban már kifejtett adatgyűjtés (PDB, Uniprot, Pfam etc.).
- Az adatok felhasználásával a fehérje gyöngyosorrepresentációjának elkészítése következik. Ez a reprezentáció nagyon fontos lépés, mert meghatározza azokat az alap tényezőket, amelyek befolyásolják a későbbi komplex összerakásának lehetőségeit (lásd 4.5 ábrán).
- Ki kell választani a szimulációs folyamat során használandó megszorításokat (Exclude Volume, Cross-links).
- A megszorítások ismeretében már elindítható a fehérjemodellek egymáshoz illesztésére a Monte Carlo szimuláció.
- A legvégén, mielőtt kiválasztjuk a legalkalmasabb modelleket, részletes elemzésre van szükség (szűrési lépések, klaszterezés, távolságok és lokális sűrűségek vizsgálata).

Fontos ellenőrizni a szerkezeti modellek megbízhatóságát, és ehhez kísérleti adatokra van szükség. A szerkezetek validálására jó módszer lehet a FOXS (Fast X-Ray Scattering) eljárás. A FoXS [63], [64] validátor sugárzás szórásán alapul: amikor az adott komplex oldatát röntgen megvilágítja akkor a komplex szét fogja szórni a röntgen sugarakat. Ezeket a sugarakat egy detektor érzékeli, majd ezen elszórt jelekből tudunk egy görbét létrehozni, amely a szórási görbe lesz - a kísérletes eljárás neve a SAXS (Small Angle X-ray Scattering). Hasonló szórási görbét matematikai úton is tudunk számolni 3D szerkezetből - ez a módszer a FoXS, így a kísérleti szórásgörbét rá tudjuk illeszteni a térszerkezet modelljeiből keletkezett szórásgörbékre ezáltal az illesztés hibája megmondja, hogy körülbelül hol s mennyire térhet el a modellünk a kísérletes komplex szerkezettől.



4.5. ábra. Az egyes fehérjék gyöngysormmodellje és a Pfam által meghatározott doménszerkezete. Itt bemutatom egy kisebb rendszer három fehérjéjének a gyöngylánc-modelljét. Minden fehérje két különböző típusú régióból áll: rugalmas és rugalmatlan részekből. A hajlékony részeknek nincs rögzített szerkezete, több és kisebb gyöngy szimulálja őket, és mozoghatnak, miközben a hajlíthatatlan résznek rögzített szerkezete van, így nagyobb és mozgathatatlan gyöngyök modellezik. Az adott típusú szakaszon belül a hajlíthatatlan esetén a gyöngyök egymáshoz képest nem mozdulhatnak el, míg a hajlékony szakasz során a gyöngyök egymáshoz képest elmozdulnak.

A SAXS adatok használatának nehézségei a szerkezet teljes ellenőrzéséhez az, hogy a kísérleti adatok nagyon szűkösek - a Small Angle Scattering Biological Data Bank [65] csak 9 lekérdezést talált a PSD-95-re, és mindegyik csak csak a PDZ1-2 tandem tartományt reprezentálta.

4.6 A posztzinaptikus denzitás fehérjéi

A modellezendő fehérjemolekulák jellemzőinek ismerete nélkülözhetetlen a modellezési eljárásokhoz. Az ágens-alapú modellekhez tudnunk kell az egyes partnerkötőhelyek jellegét és számát, valamint lehetőség szerint az egyes kölcsönhatásokra jellemző asszociációs és disszociációs rátákat. A háromdimenziós modellezéshez a megfelelő kísérletes adatok azaz a térszerkezetek rendezése és felderítése szükséges, hiányos fehérje szerkezetek esetén alapszintű fehérje modellek keresésére és/vagy doménekenkénti összerakására van szükség ahhoz, hogy őket később komplexben tudjam megjeleníteni.

Az előzőekben említettem, hogy a posztzinaptikus denzitás fehérjéinek terét hét fő fehérjére szűkítettem le. Ennek oka, hogy a szimulációk és a modellezések során a fehérjék teljes tárházát nagyon nehéz egységben vizsgálni, és ahhoz, hogy megítélhessem, hogy az én általam létrehozott pipeline segítségével vizsgálhatóak-e az általam feltett kérdések, nem

(feltétlenül) szükséges a posztszinaptikus denzitás több ezer fehérjéjét egyszerre tekinteni. Természetesen a későbbiekben lehetőség nyílik a fehérjetér bővítésére. A kiválasztott fehérjék mind modulálják a tanulás és memória hatásfokát és folyamatát [66].

Ezen hét fehérje esetében is kevés 3D szerkezeti modell áll rendelkezésünkre. Mind a hét fehérje viszonylag nagy fehérjének számít, mert a posztszinaptikus denzitásban jellemzően nagyobb fehérjék találhatóak meg a szervezet egyéb átlagos fehérjeihez képest. Többek között ezért is tűnik a szinapszis ezen területe sűrűnek az elektronmikroszkópos képeken, nem csak a fehérjék mennyisége miatt. Ezért összegyűjtöttem a meglévő szerkezeteket, amelyeket könnyedén megtaláltam az rcsb.org-on keresztül és megvizsgáltam, hogy mely szerkezetek között találok átfedéseket illetve milyen fehérje részek vannak, amelyekre nem találtam szerkezetet, hiszen a nagy méret miatt javarészt doménenként tudták a szerkezeteket meghatározni, a globuláris részekre koncentrálnak. Szerencsére a kutatócsoport, amellyel együtt dolgozom ezen fehérjék szerkezetét vizsgálja kísérleti és elméleti módszerekkel is (például a Shank1 PDZ doménjének szerkezetét [67]).

A IMP modellező program a fehérje dinamikusan mozgó laza szakaszait viszonylag dinamikusan kezeli, viszonylag nagy felbontásban, részletgazdagabban, míg az együtt mozgó domén régiókat egy-egy nagyobb együtt mozgó, gömbökből felépített merev egységként reprezentálja. Ezért lehetőségem nyílik arra, hogy a különböző térszerkezetekből származó doméneket flexibilis összekötő (linker) régiók segítségével összeilleszthessem, amennyiben a domének közötti hiányzó fehérje részek dinamikus mozgást mutatnak.

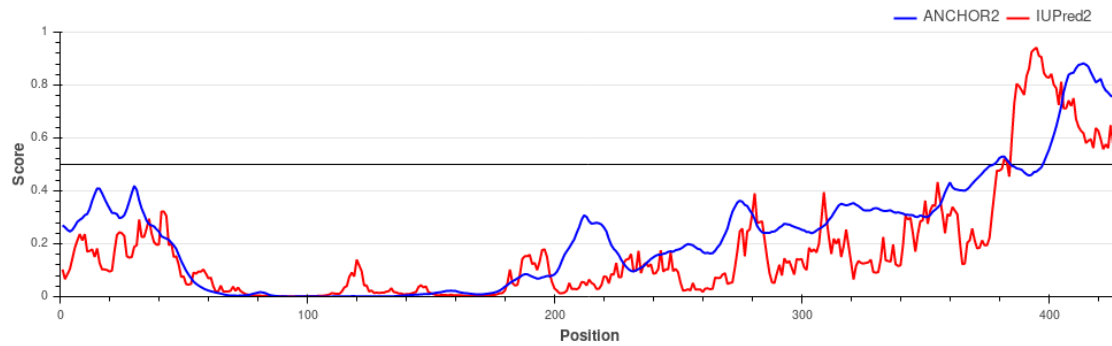
4.6.1 A Kir2.1 káliumion-csatorna

A Kir2.1 az egyik legfontosabb káliumion-csatorna. A csatorna a funkciója révén kapta a nevét: (inwardly rectifying potassium) befeléáramlással helyreállító káliumcsatorna, ami lehetővé teszi a kálium áramlását az intracelluláris mátrixba, és alig enged visszafelé áramot. A Kir2.1 csatornák elhelyezkedése befolyásolhatja az idegsejt működését, mivel a Kir2.1 csatornák túlexpressziója hiperpolarizációhoz vezethet a káliumáramok miatt. A fehérje négy 427 aminosavból álló alegység heterotetramere [68]. Az IUPred2A [69] szerint a Kir2.1 rendezett fehérje kis rendezetlen C-terminális szegmensekkel, amint az látható a 4.6. ábrán. A C-terminális PDZ-domént kötő motívum a Glu-Ser-Glu-Ile szekvenciával (residue 423–427) - pont ez a szakasz tűnik a legrendezetlenebbnek, amelyen keresztül a Kir2.1 PSD-95-tel koprecipitálódik [70].

A Pfam adatbázis szerint a fehérje két fő doménre osztható az N-terminális szakasz után: Káliumion-áramlást befelé egyenirányító káliumcsatorna transzmembrán doménre - IRK (residue:48-186) és egy C-terminális doménre - IRK_C (residue: 193 -366), a két domén között transzmembrán régióval A C-terminálison pedig a már említett rendezetlen régióval.

4.6.2 Az NMDA receptor

Az NMDA receptor elengedhetetlen a hosszú távú potencióció kiváltásához, amely tartós változás az idegsejtek reakciókészségében a tanulás és az emlékezet alapjául szolgál [71]. Az NMDA receptor szerepe a hosszú távú erősítés kiváltása kapcsán abban rejlik, hogy ahogy Donald Hebb 1949-ben felismerte, koincidencia-detektorként működik - akár egy ÉS kapu - mivel feszültségfüggő Mg^{2+} kötőhelyei blokkolják a kationcsatornát depolarizáció hiányában. Tehát a csatorna akkor és csak akkor van nyitva, ha egyszerre köt ligandum hozzá és lokális feszültségváltozás van. A receptor Ca^{2+} permeabilitása és másodlagos hírvivő útvonala



4.6. ábra. A Kir2.1 fehérje IUPred2A által jóslott rendezett-rendezetlen régiói. A piros és kék görbék két különböző algoritmus alapján jósolják meg a rendezettség mértékét. A fekete vízszintes vonal a küszöbérték, amely fölött rendezetlennek tekinthető a fehérje.

[72] révén az egyik legfontosabb komponens, amely a hosszú távú erősödését befolyásolja. Szükséges a fel nem használt szinapszisok hatékony eliminálásához is a szinaptogenezis második fázisában [73].

A receptor alegységei közül az NMDAR1 alegység viszonylag egységesen expresszálódik a neuronokban, míg az NMDAR2 különböző típusai átlapolással, de különböző mennyiségben expresszálódnak, így egyfajta NMDA receptoron belüli diverzitást létrehozva, amely kihat a szinapszis tulajdonságaira [74]. A diverzitást tovább növeli, hogy noha az első alegységből általában kettő azonos, míg a második alegységből akár két különböző típus is kerülhet egyetlen heteromer csatornába [75].

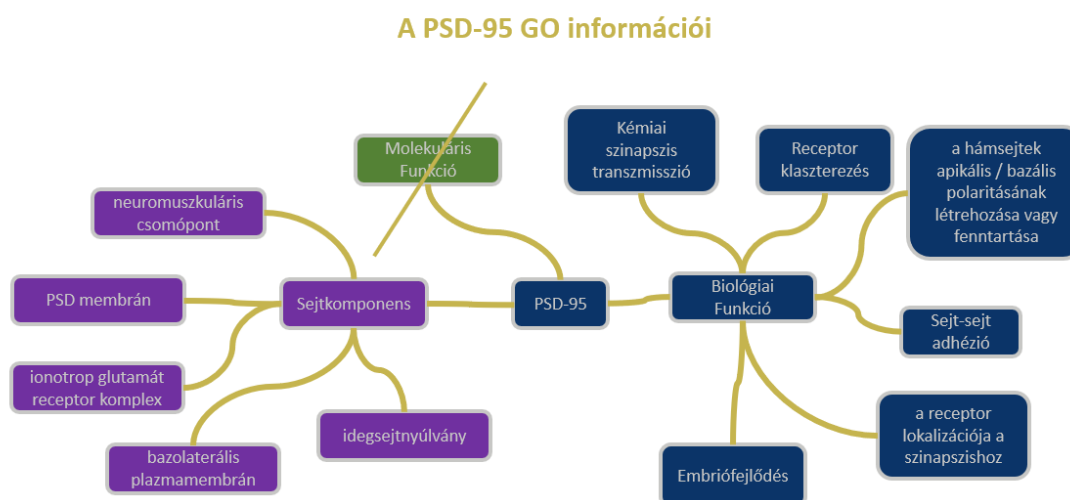
A Pfam szerint négy fő domént találunk, amelyek a következők: ANF-receptor (39-360 aminosavak) [76], majd a rendezetlen összekötő régió után az L-glutamát- és glicinkötő régió (residue 442-546) [75], [77] majd egy kisebb rendezetlen régió után az ionotróp glutamát receptor (residue 559-824), majd a transzmembrán régió után találjuk a Calmodulin-kötő domént (residue 835-863) [78].

Az NMDA receptorok PSD-95 hiányában is képesek a megfelelő lokalizációra [79], az ezt leíró tanulmány azonban a jelátviteli útvonalak teljes működését nem vizsgálja. Ettől függetlenül mai ismereteink szerint a PSD-95 elengedhetetlen a PSD háromdimenziós molekuláris szerkezetének fenntartásához [80].

4.6.3 Az AMPA receptor

Az AMPA receptor szintén gyakori ionotróp glutamát receptor, amely az NMDA receptorhoz hasonló felépítés, ám különböző alegységekkel rendelkezik, ami eltérő Ca^{2+} permeabilitást és a feszültségfüggő viselkedés hiányát eredményezi [81]. Mind az NMDAR, mind az AMPAR szükséges a glutamáterg szinapszisok normális működéséhez.

Az NMDA receptorokkal ellentétben az AMPA receptorok esetében a PSD-95 fehérje nélkülözhetetlen ahhoz, hogy megfelelő mennyiségben lokalizálódni tudjanak a poszt-szinaptikus denzitás periferiáján [80]. Az AMPAR képes a sejt felszínén oldalirányban elmozdulni vagy változtatni helyzetét a membránon belül. Az AMPAR az idegsejtek membránjában történő változása lehetővé teszi az idegsejtek számára, hogy gyorsan alkalmazkodjanak a környezeti változásokhoz és az idegi aktivitáshoz. LTP során a szinapszis középpontja felé vándorolnak, míg depresszió során a centrumtól távolodó irányba növekszik meg a vándorlás mértéke [82].



4.7. ábra. **A PSD-95 biológiai funkciói és előfordulásai** A PSD-95 egy fontos fehérje a posztszinaptikus sűrűségben, amely részt vesz a kémiai szinapszisok transzmissziójában és a receptorok klaszterezésében. A biológiai funkciói közé tartozik a sejt-sejt adhézió, a hámsejtek apikális/bazális polaritásának létrehozása és fenntartása, az embriófejlődés, valamint a receptorok szinapszisokhoz történő lokalizálása. A sejtkomponens szinten a PSD-95 jelen van a neuromuskuláris csomópontban, a PSD membránban, az ionotrop glutamát receptor komplexben, a bazolaterális plazmamembránban és az idegsejtnyúlványokban. Ez az ábra tehát átfogó képet ad a PSD-95 szerepéről és előfordulásáról a sejtekben és a különböző biológiai folyamatokban. Az ábra a szerző saját munkája.

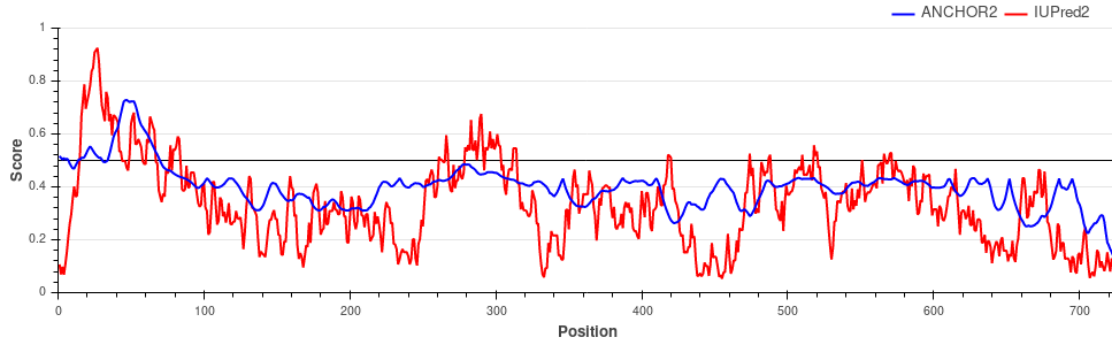
Az AMPAR laterális diffúziója is hozzájárulhat a posztszinaptikus denzitásban megfigyelhető fázisszeparációs folyamatokhoz, vagy fordítva. Szuper-rezolúciós mikroszkópiával az AMPAR nanodoménekké rendeződése megfigyelhető. Egy adott nanodoménen belül az AMPAR elveszti mobilitását, a nanodoménen kívül a laterális vándorlás lényegesebb. Az AMPAR mobilitása nem csak pozíciójától függ, hanem a szinapszis aktivitásától is - aktív szinapszis esetén az AMPAR pozíciója stabilizálódni látszik az inaktív szinapszishoz képest [83]–[85]. Ez a folyamat is az LTP stabilizáló, szinapszis-megerősítő hatását mutatja .

4.6.4 A PSD-95 fehérje

A PSD-95 (posztszinaptikus denzitás fehérje, 95 kDa) az egyik leggyakoribb állványfehérje a posztszinaptikus denzitásban, ezért is úgy tartják, hogy főszervező szerepet tölt be a PSD-ben. A posztszinaptikus denzitás fehérjei nem képeznek strukturált hálózatot a PSD-95 hiányában [80]. A Gene Ontology alapján a PSD-95 szerepet játszik (lásd 4.7. ábra) a Huntington-kór kialakulásában, hiszen ezen a fehérjén keresztül befolyásolja a kalcium-csatornákat, azáltal, hogy a Huntingtin a PSD-95-tel asszociál [86].

A PSD-95 az ún. MAGUK (membrane-associated guanylate kinase) családba tartozik, három PDZ-domént tartalmaz, egy SH3, valamint egy GK domént. Az N-terminálisa a

leginkább rendezetlen régió [87] - lásd 4.8. ábra. Ez az információ funkcionálisan értelmes, mert az N-terminálisok a PSD-95 palmitoilációjának vannak kitéve. Az N-terminális rugalmassága segíti a fehérje ilyen módon való lehorgonyzását a membránban [88].



4.8. ábra. A PSD-95 fehérje IUPred2A által jósolt rendezett-rendezetlen régiói. A piros és kék görbék két különböző algoritmus alapján jósolják meg a rendezettség mértékét. A fekete vízszintes vonal a küszöbérték, amely fölött rendezetlennek tekinthető a fehérje. Az ábra a szerző saját munkája

A PSD-95 PDZ3 doménjén keresztül is képes microtubulusokhoz kapcsolódni közvetve a CRIPT névre hallgató mediátorral, rögzítve a PSD-t a citoskeletonhoz [79].

4.6.5 A SynGAP fehérje

A SynGAP-re azért esett a választásom, mert jelentős hatást gyakorol az agyban, többek között úgy, hogy izoformái egyedi spatiotemporális expressziós mintázatot mutatnak, és különböző szerepet játszanak a neuronális és szinaptikus fejlődésben [89]. A SynGAP egy olyan fehérje, amely aktiválja a Ras és Rap kis G-fehérjéket, segítve azokhoz kötött GTP (aktív) átalakítását GDP-re (inaktív), ezzel negatívan szabályozva a kis G-fehérjék aktivitását [89], [90]. Azáltal, hogy a RAS fehérjét gátolja, teljes jelátviteli kaszkádokat befolyásol (MAPK, IP3K etc.), amelyeket a RAS képes beindítani. Ez a kaszkádinhibíció egyes fehérjék expressziójának csökkenését eredményezi, így a teljes posztzinaptikus denzitás szerkezetét is megváltoztatja - többek között a posztzinaptikus membránba ágyazott AMPAR mennyiségét csökkenti [91].

Azokban a hippocampus CA1 piramis sejtekben, amelyekben a SynGAP fehérje hiányzik, a hosszú távú potenciáció (LTP) rendkívül súlyosan károsodik [89], [92]. Ezen piramis sejtek kritikusak az emlékezeti folyamatokban és a tanulásban, így az LTP hibás működése jelentős hatással van a memória képességekre és az információ feldolgozására, így a munkamemóriára is [89], [93].

4.6.6 A GKAP fehérje

A GKAP C-terminális szegmense kölcsönhatásba léphet a Shank3 PDZ doménjével. Hét fehérjéből álló készletünkben a GKAP tekinthető kapcsolatnak a PSD-95 által potenciálisan összekötött receptorok rétege és a homooligomerizációra képes Shank3 és Homer1, mélyebb rétegre jellemző állványfehérjék között [94].

A szabad GKAP elérhetősége függ a szinapszis aktivitásától, mert az aktivitás során beáramló kalcium-ion felszabadítja az aktinfonalakon tárolt GKAP-ot, azáltal, hogy az

aktivált CAMKII β foszforilálja a Ser340 és Ser384 aminosavakat, így a GKAP leválik az aktinról [94]. Ezen függőség által az LTP alatt bekövetkező szinapszis megerősítésben a GKAP lehetővé teszi, hogy a receptor/PSD-95/SynGAP komplex pozíciója stabilizálódjon aktivitás hatására. A GKAP hiánya egyenesen szinapszis megszűnéshez is vezethet [94].

A GKAP állványfehérje szinte teljesen belsőleg rendezetlen, emiatt kevés a róla elérhető térszerkezeti információ. Több kötőhelyet tartalmaz a GK doménhez és a DYNLL-hez, és van egy helikális GH1 doménje a C-terminális közelében. [95] A GKAP foszforilációtól függő módon képes a PSD-95 GK doménjével kölcsönhatásba lépni.

4.6.7 A Shank3 fehérje

A Shank3 fehérje fő szerepet játszik abban, hogy összeköti az NMDA receptorokat a dendrit oldalsó, metabotróp glutamát receptorokkal [96], ezekhez a PDZ doménjén keresztül kapcsolódik, hasonlóan, mint a PSD-95 az NMDA receptorokhoz.

A Shank3 a Shank (SH3 és több ankyrin ismétlődő domén fehérje) család tagja, amely egy ankyrin ismétlődő régiót és egy SH3, egy PDZ és egy SAM domént, valamint egy prolinban gazdag szegmenst ([97]) tartalmaz. A Shank3 jól ismert az autizmus spektrumzavarban betöltött szerepéről, és túlzott kifejeződése az ADHD-vel és a szinaptikus diszfunkcióval függ össze [98][99].

A Shank3 is részt vesz az LTP folyamatában. Autizmus spektrum zavar tud fellépni a cink nem megfelelő mennyisége hatására. A kapocs a cink mennyisége és a szinapszis működése között a Shank3 fehérje. A cink a SAM doménon keresztül aktiválja a Shank3 fehérjét és elérhetővé teszi azt a posztzinaptikus denzitás komplexei számára [100] - a GKAP-hoz hasonlóan megerősödik a szinapszis ezáltal (csak ebben az esetben a cink az aktivátor molekula).

A Shank3 a SAM doménjén keresztüli önasszociációra képes [101]. A Shank3 az a fehérje, amely a PSD-95:Homer1-komplezset felelős. PDZ doménje képes megkötni a GKAP C-terminálisát, a prolingazdag régiót pedig a Homer1 [96] EVH1 doménje ismeri fel.

4.6.8 A Shank1 fehérje

A Shank1 is a Shank család tagja, hasonlóan viselkedik a Shank3-hoz.

A Shank fehérjecsalád nagy moduláris állványfehérjéket tartalmaz, globuláris és hosszú rendezetlen régiókkal [102]. Ezek a fehérjék számos változatos kölcsönhatást hoznak létre különféle posztzinaptikus fehérjékkel. A Shank fehérjék hozzáférhetőségét és/vagy szerkezetét befolyásoló mutációkat számos állapottal hozták összefüggésbe [103] az autizmus spektrum zavartól (ASD) [104] a Phelan-McDermid szindrómáig (PMS) [105]. Ilyen körülmények között a Shank3 haploid elégtelenség (haploinsufficiency) [106], amelyet vagy a gén egy másolatának teljes elvesztése vagy egy funkciót befolyásoló mutáció jelenléte okoz, a leggyakoribb ok. A Shank1 és Shank2 gének mutációi hasonló, bár jellemzően enyhébb fenotípusokat is okoznak.

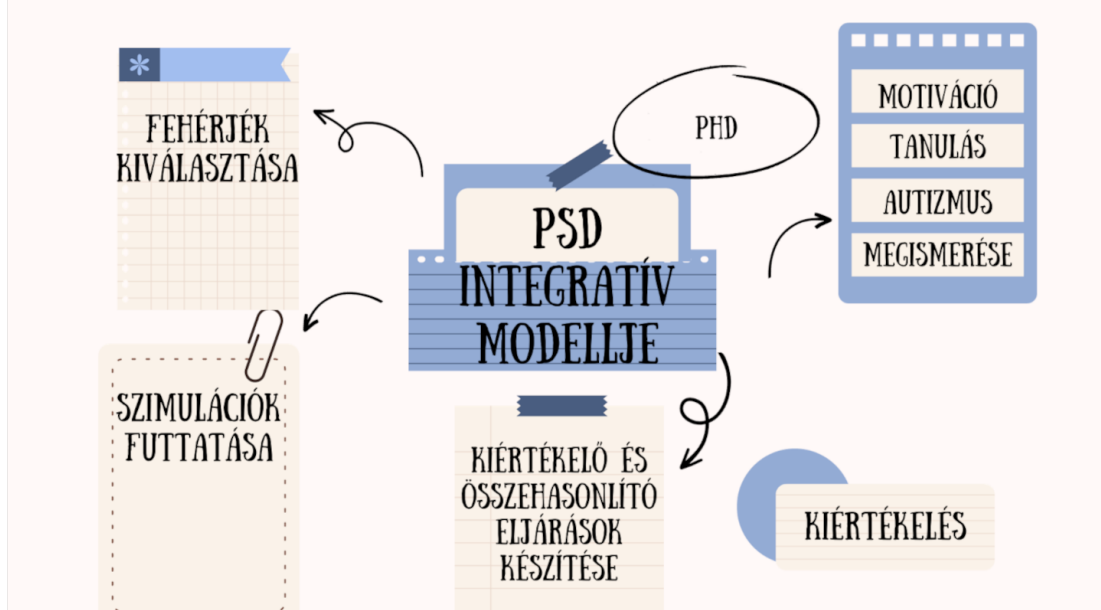
Szerkezetében eltér, de domének szintjén ugyanúgy az ankyrin ismétlődő régióval, az SH3, PDZ, valamint SAM doménekkal rendelkezik a megfelelő sorrendben [107]. A legnagyobb eltérés a domének abszolút helyzetében jelenik meg. Például a legelső ANK régió a 148 helyett a 212. aminosavnál kezdődik. A domének eltérő pozíciója a relatív mozgásukat befolyásoló lényeges tényező, amely a 3D modellezéseknél játszhat szerepet.

4.6.9 A Homer1 fehérje

A Homer1 egy olyan vázfehérje, amely szabályozza a glutamáterg szinapszisokat és a dendrittüskék morfogenezisét [108]. Az EVH1 domainjén keresztül az aktinkötegek rendeződését szabályozza [109]. Az AMPAR újrahasonosítását és a metabotróp glutamát receptorok elrendeződését/hatását is szabályozza [110]–[112]. Ez a szabályozás többek között elhelyezkedéséből adódik a PSD-ben. Amennyiben a Homer1 nem elérhető, foszforilációs folyamatok miatt, a membránban elhelyezkedő metabotróp glutamát receptorok hatása is csökken. Tehát a Homer1 kapcsolóként működik, amely ki/bekapcsolásával a receptorok hatása modulálható [112].

Az EVH1 egy hosszú tekerces coiled-coil szegmens, amely a Homer1 molekulák homodimerizációját és tetramerizációját is közvetíti. Az EVH1 doménje továbbá kötődik a Shank3-hoz [96]. Az EVH1 domén prolinban gazdag szegmenseket képes megkötni Shank fehérjéken, metabotrop glutamát receptorokon, valamint IP3 és rianodin receptorokon [113].

Több izoformja is létezik, a Homer1a izoformából hiányzik a tekerces tekercs régió, és szerepet játszik a PSD átszervezésében alvás közben [114]. A Homer1a izoform expressziója a hosszabb izoformokhoz kötés valószínűségét csökkenti, amellyel inhibálja a metabotróp glutamát receptorok moduláló hatását, ami kihat az NMDA és AMPA receptorok által létrejövő LTP mértékére és gátolja azt - átrendi a teljes posztzinaptikus denzitást [115].

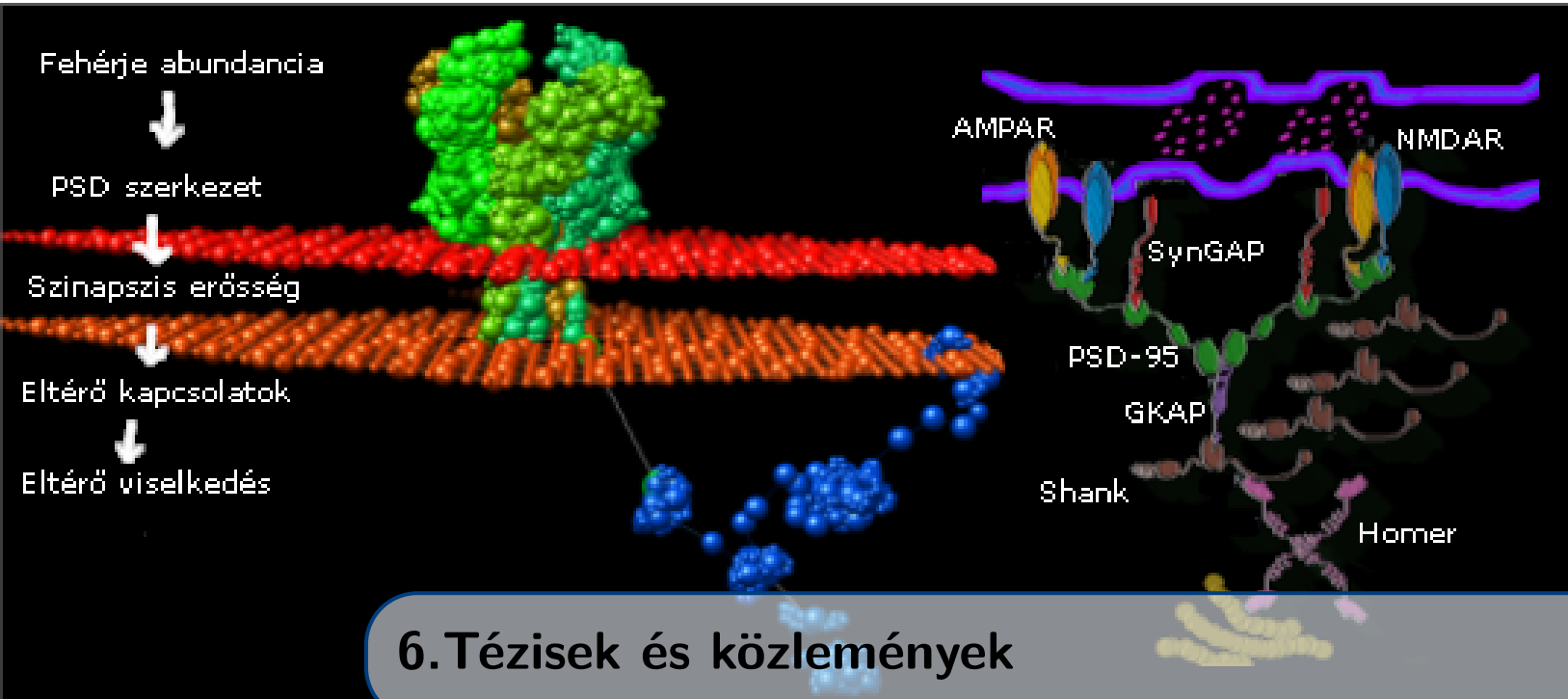


5. Célkitűzések

Kutatásom fő célja a posztszinaptikus denzitás integrált modellezéséhez szükséges alapvető eljárások kidolgozása és alkalmazása volt. Terveim szerint az integrált modellezés segítségével az egyes fehérjekomplexek kialakulásának valószínűségein túl szerkezeti modelleket is készíthetünk, figyelembe véve az egyszerűsített szimulációkból kapott információkat. Ennek megvalósításához az alábbi lépések megvalósítását tűztem ki célul:

- Az igen sok fehérje összetett interakciója által felépülő posztszinaptikus denzitás egyszerűsített, kezelhető méretű modelljének kialakítása, melynek segítségével ágens-alapú szimulációk végezhetőek. Ehhez szükséges a leggyakoribb fehérjék, azok példányszámának és egymással kialakítható kölcsönhatásainak azonosítása és reprezentációja oly módon, hogy az a tervezett szimulációk bemeneteként szolgálhasson.
 - *Miért?* A PSD rendkívül komplex és nagyszámú fehérje kölcsönhatásából épül fel. Egy egyszerűsített, kezelhető méretű modell lehetővé teszi az ágens-alapú szimulációk elvégzését, amelyek segítségével megérthetjük az alapvető mechanizmusokat és interakciókat a PSD-ben.
 - *Miért fontos?* Az ágens-alapú szimulációk lehetővé teszik a PSD dinamikus viselkedésének és a fehérjekomplexek kialakulásának vizsgálatát, ami kritikus a funkcionális megértéshez és a terápiás beavatkozások fejlesztéséhez.
- A modell segítségével elvégzett ágens-alapú szimulációk kiértékeléséhez szükséges eljárások kidolgozása, ideértve a létrejött fehérjekomplexek egyértelmű azonosítását, illetve különböző bemeneti paraméterekkel indított szimulációk összehasonlítását.
 - *Miért?* A szimulációk eredményeinek pontos értékelése kulcsfontosságú a modellek validálása és a predikciók megbízhatóságának biztosítása érdekében.
 - *Miért fontos?* Az eljárások kidolgozása biztosítja, hogy a szimulációk eredményei összehasonlíthatóak és reprodukálhatóak legyenek, lehetővé téve a különböző hipotézisek tesztelését és a biológiai relevancia megértését. A feladat nem triviális többek között azért, mert a lehetséges komplexek száma elméletben végtelen. Adott komplexek eltérése lehet elenyésző, vagy pont kimagasló funkcionalitása miatt.

- Nagyszámú szimuláció elvégzése számos különböző bemeneti paraméter felhasználásával irodalmi fehérjegyakorisági adatok segítségével, majd a szimulációk kiértékelése és értelmezése.
 - *Miért?* A különböző bemeneti paraméterek tesztelése segít megérteni a PSD kialakulásának és működésének variabilitását, valamint az egyes fehérjekomplexek szerepét a rendszer stabilitásában és dinamikájában.
 - *Miért fontos?* A széleskörű szimulációk lehetővé teszik a rendszer robusztusságának és érzékenységének vizsgálatát, ami fontos a biológiai rendszerek pontos modellezéséhez és a potenciális gyógyszer célpontok azonosításához.
- Adott kölcsönhatást gyengítő mutáció hatásának modellezése a keretrendszer használatával.
 - *Miért?* A mutációk hatásának modellezése segít megérteni a genetikai variációk és mutációk hatását a PSD működésére és stabilitására.
 - *Miért fontos?* Az ilyen modellezések segíthetnek azonosítani a potenciális patogén mutációkat és megérteni azok mechanizmusait, ami fontos a genetikai betegségek diagnosztikájában és kezelésében.
- Mindezekkel párhuzamosan munkafolyamat kidolgozása és tesztelése nagyméretű, többkomponensű posztszinaptikus fehérjekomplexek háromdimenziós modellezéséhez.
 - *Miért?* A nagy és komplex fehérjekomplexek térbeli modellezése kihívást jelent, de elengedhetetlen a teljes PSD szerkezeti és funkcionális megértéséhez, hiszen önmagában az ágens alapuló szimulációk csak gráfként tekintenek a komplexekre és éppen ezért a térbeli gátlások és a fehérjék/komplexek térigénye nem jelenik meg.
 - *Miért fontos?* Az ilyen modellek lehetővé teszik a szinaptikus funkciók és a fehérje-interakciók finom mechanizmusainak vizsgálatát, amelyben már térbeli, szerkezetbeli hatások is észlelhetők hozzájárulva a teljes idegrendszeri működés megértéséhez és a potenciális terápiás beavatkozások fejlesztéséhez.

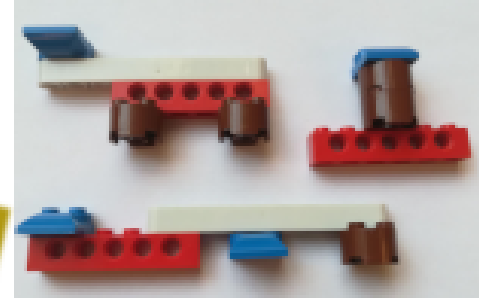
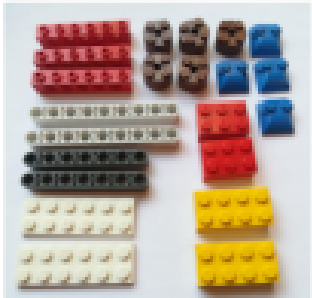


6. Tézisek és közlemények

1. Irodalmi adatok alapján előkészítettem a posztszinaptikus denzitás (PSD) fehérje-hálózatának leegyszerűsített modelljét, mely 7 fő fehérjét tartalmaz és bemenetként használható fehérjekomplexek eloszlását szimuláló rendszerbiológiai számításokhoz. Expressziós adatok alapján a hét fehérjét összesen 524 különböző gyakorisági összetételben vizsgáltam a Cytocast szoftver segítségével. (Miski et al. 2022, [1])
2. Eljárásokat dolgoztam ki a PSD modell szimulációiból kapott fehérjekomplex-eloszlások összehasonlító elemzésére. Az eljárások lehetővé teszik az egyes fehérjekomplexek egyértelmű azonosítását és ennek segítségével a különböző szimulációs eredmények összevetését eltérő kiindulási fehérjegyakoriságok és a partnerkötési tulajdonságokat befolyásoló mutációk esetében is. (Miski et al. 2022, [1])
3. Megállapítottam, hogy a keletkező fehérjekomplexek eloszlása és az egyedi fehérjék gyakorisága közötti kapcsolat összetett, csak szimulációkkal térképezhető fel. Nagyon hasonló kiindulási fehérjegyakoriságok is eredményezhetnek jelentősen eltérő fehérjekomplex-eloszlást, melynek a szinaptikus elmélet alapján funkcionális jelentősége lehet. (Miski et al. 2022, Miski proc. 2020 [1], [2])
4. Megmutattam, hogy egy hipomorf, azaz egy adott fehérje-fehérje kölcsönhatást csak kismértékben gyengítő mutáció hatása jelentősen függ az adott sejtekre jellemző fehérjegyakoriságoktól, emiatt csak adott sejtekben vagy sejtípusokban okoz szignifikáns változást. Mindemellett az is jellemző, hogy a legnagyobb változást mutató komplexek nem feltétlenül azok, amelyekben változással érintett interakciós partnerek szerepelnek. (Miski et al. 2024, [3])
5. Kidolgoztam egy munkafolyamatot a posztszinaptikus denzitásra jellemző egyes nagyméretű, többkomponensű fehérjekomplexek lehetséges háromdimenziós szerkezeteinek modellezésére. (Miski proc. 2019,2020,2022, [4]–[6])

Az értekezés alapjául szolgáló közlemények

- [1] M. Miski, B. M. Keömley-Horváth, D. R. Megyeriné, A. Csikász-Nagy és Z. Gáspári, “Diversity of synaptic protein complexes as a function of the abundance of their constituent proteins: A modeling approach”, *PLOS Computational Biology*, 18. évfolyam, 1. szám, M. Migliore, szerkesztő, e1009758., 2022. jan. DOI: [10.1371/journal.pcbi.1009758](https://doi.org/10.1371/journal.pcbi.1009758). cím: <https://doi.org/10.1371/journal.pcbi.1009758> (hivatkozási oldal **33**).
- [2] M. Miski és A. Csikász-Nagy, “Analyses of protein-protein interactions in the PSD by stochastic simulations”, *PHD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY AND BIONICS*, 15. évfolyam, 88–90. oldal, 2020, ISSN: 2064-7271 (hivatkozási oldal **33**).
- [3] M. Miski, Á. Weber, K. Fekete-Molnár, B. M. Keömley-Horváth, A. Csikász-Nagy és Z. Gáspári, “Simulated complexes formed from a set of postsynaptic proteins suggest a localised effect of a hypomorphic Shank mutation”, *BMC Neuroscience*, 25. évfolyam, 1. szám, 2024. júl., ISSN: 1471-2202. DOI: [10.1186/s12868-024-00880-1](https://doi.org/10.1186/s12868-024-00880-1). cím: <http://dx.doi.org/10.1186/s12868-024-00880-1> (hivatkozási oldal **33**).
- [4] M. Miski, K. Kornél, A. Csikász-Nagy és Z. Gáspári, “Integrative modeling of possible layouts of Kir2.1 localized in membranes and connected by PSD-95”, *JEDLIK LABORATORIES REPORTS*, VII. évfolyam, 29–32. oldal, 2019, ISSN: 2064-3942 (hivatkozási oldal **33**).
- [5] M. Miski, “Analyses of protein-protein interactions related to nNOS by stochastic simulations”, *PHD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY AND BIONICS*, vol. 14 évfolyam, 59–62. oldal, 2019, ISSN: 2064-7271 (hivatkozási oldal **33**).
- [6] M. Miski és A. Csikász-Nagy, “Distribution and structure of postsynaptic protein complexes assessed by simulations”, *PHD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY AND BIONICS*, vol. 16 évfolyam, 50. oldal, 2021, ISSN: 2064-7271 (hivatkozási oldal **33**).



7. Módszerek

A Cytocast, SiCompre, IMP, a saját oldalaiikon elérhetőek. A saját postprocessing kódomat a szimulációhoz feltöltöttem. [GitHub projektem](#).

7.1 Fehérjeadatbázisok

A fehérjeadatbázisok rendkívül fontosak a tudományos kutatásom szempontjából. Az ilyen adatbázisok segítségével rendelkezésre állnak az ismert fehérjék struktúrái és funkciói, fontos tulajdonságai, aminek köszönhetően azonosíthatom hasonlóságait és evolúciós kapcsolatait. Az adott fehérjékhez tartozó megfelelő gének ismerete is nélkülözhetetlen. Az alfejezetben szeretném bemutatni azokat a főbb adatbázisokat, melyeket felhasználtam a modell összerakása során.

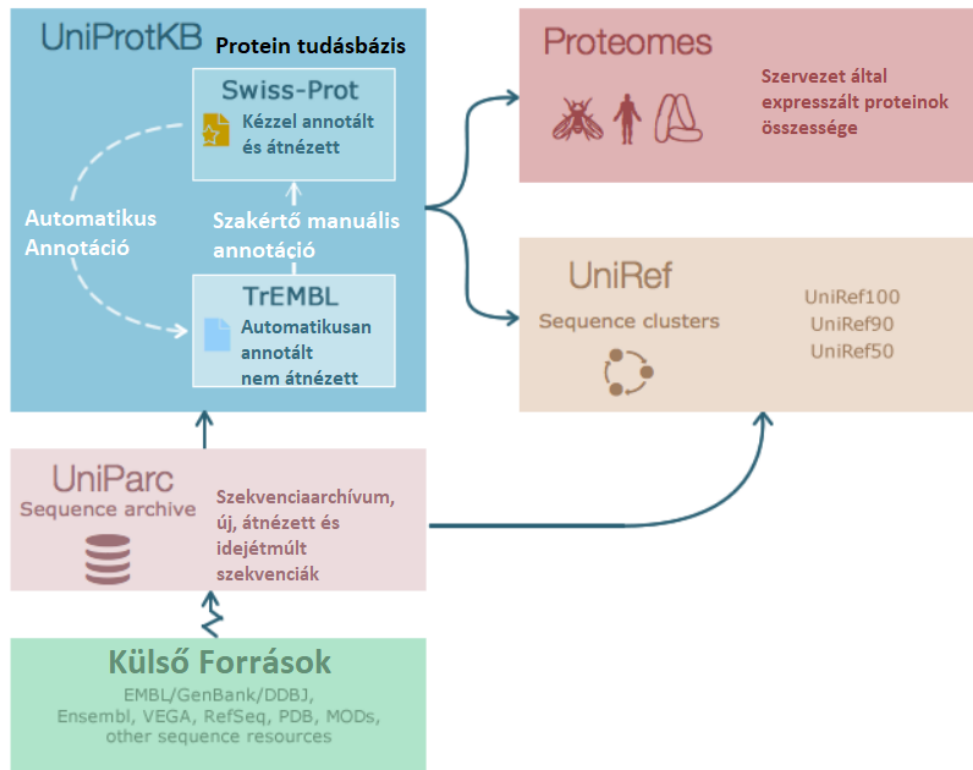
7.1.1 UniProt

Az UniProt [116] küldetése, hogy "a tudományos közösség számára átfogó, jó minőségű és szabadon hozzáférhető fehérjeszekvencia- és funkcionális információforrást biztosítson."

Az UniProt adatbázist a fehérjék azonosítására és azok annotációjára használtam. Ez magában foglalja a fehérjék aminosav szekvenciáinak meghatározását, valamint a hozzájuk kapcsolódó biológiai funkciók, molekuláris interakciók és celluláris lokalizációk részletes leírását. Az UniProt adatbázisban található adatok segítettek azonosítani azokat a fehérjéket, amelyek kulcsszerepet játszanak a poszt-szinaptikus denzitás (PSD) felépítésében és működésében.

Az UniProt adatbázisból származó információk felhasználásával részletes elemzést végeztem a fehérjék funkcióiról és azok kölcsönhatásairól más molekulákkal. Ez különösen fontos volt a komplex fehérje-interakciós hálózatok megértéséhez, amelyek alapvetőek a PSD stabilitása és dinamikája szempontjából. Az UniProt adatbázis lehetővé tette számomra, hogy azonosítsam a leggyakrabban előforduló fehérjéket és azok interakciós partnereit.

Az UniProt három fő adatbázisra különül el: UniProt Knowledgebase (UniProtKB), UniProt Reference Clusters (UniRef), az UniProt Archive (UniParc) (lásd 7.1. ábra).



7.1. ábra. **Az Uniprot adatbázis működése** Ez az ábra a UniProtKB fehérje adatbázis szerkezetét és működését mutatja be. A UniProtKB két fő részre oszlik: Swiss-Prot és TrEMBL. A Swiss-Prot manuálisan annotált és ellenőrzött adatokat tartalmaz, míg a TrEMBL automatikusan annotált és nem ellenőrzött adatokat. Az annotációs folyamat során a TrEMBL adatokból manuális ellenőrzéssel Swiss-Prot adatok lesznek. Az adatbázis kapcsolódik a UniParc-hoz, amely a szekvencia archívumot tartalmazza, beleértve az új, módosított és elavult szekvenciákat. Az UniParc adatokat különböző külső forrásokból, mint az EMBL, GenBank, DDBJ, Ensembl, VEGA, RefSeq, PDB és más szekvencia forrásokból származnak. Az adatbázis továbbá kapcsolatban áll a Proteomes-el, amely az organizmusok által kifejezett fehérjekészleteket tartalmazza, és a UniRef-el, amely szekvencia klasztereket tartalmaz (UniRef100, UniRef90, UniRef50). Az átdolgozott ábra eredeti forrása az [UniProt adatbázis](#).

Legfőbb különlegessége, hogy számos a szakirodalomból szakértő kurátorok által kigyűjtött annotációkat tartalmaz [116]. A tudományos publikációkból így kinyert információkat a UniProtKB/Swiss-Prot adatbázis UniProt Knowledgebase szekciója tartalmazza. Ezeket az adatokat *evidence* jelzéssel látják el, amelyről könnyen elérhetőek a forrásai, így egyszerre számos kísérlet eredményéhez férhetünk hozzá. Már több mint 227 millió szekvenciát tartalmaz [117]. A *Homo sapiens* irodalmának körülbelül 50%-t dolgozták fel [116]. Egyik legújabb bővítésük, hogy vírus referencia proteommal is bővítették az adatbázisokat.

Az automatikus annotációk klasszifikáció segítségével történnek, mely során fehérjecsaládokba osztják az adott fehérjét, vagy lehetséges doméneket keresnek például szekvencia alapján. A két predikciós rendszerük: UniRule és Statistical Automatic Annotation System (SAAS). Külön érdekesség, hogy a cikkeket is klasszifikálják neurális hálózatokkal, hogy minél több információt gyorsan és automatikusan kiszedhessenek belőlük, ezeket is általában fehérjeközpontúlag.

Azért is különleges és fontos kiemelni ezt az adatbázist, mert a fehérjék egyedi azonosítójával könnyedén beazonosítható minden fehérje és általa nem keverjük össze a fehérjét esetlegesen úgymond ugyanazon fehérje másik szervezetben lévő adataival sem. Én igyekeztem az UniProt IDk alapján megkülönböztetni a használt fehérjéket, így a fehérjék különböző nevei kevésbé okozhatnak félreértéseket.

Az innen szerzett információkat használja fel például a komplexeket összegyűjtő Complex Portal (<https://www.ebi.ac.uk/complexportal/>) is.

7.1.2 DisProt

Amennyiben rendezetlen fehérjékkel, vagy fehérjerégiókkal találkozunk, akkor az erre szakosodott DisProt (<https://disprot.org/>) adatbázis [118] segíthet az annotációk megtalálásában [119]. A DisProt annotációk felépítéséről lásd az összefoglaló 7.2. ábrát.

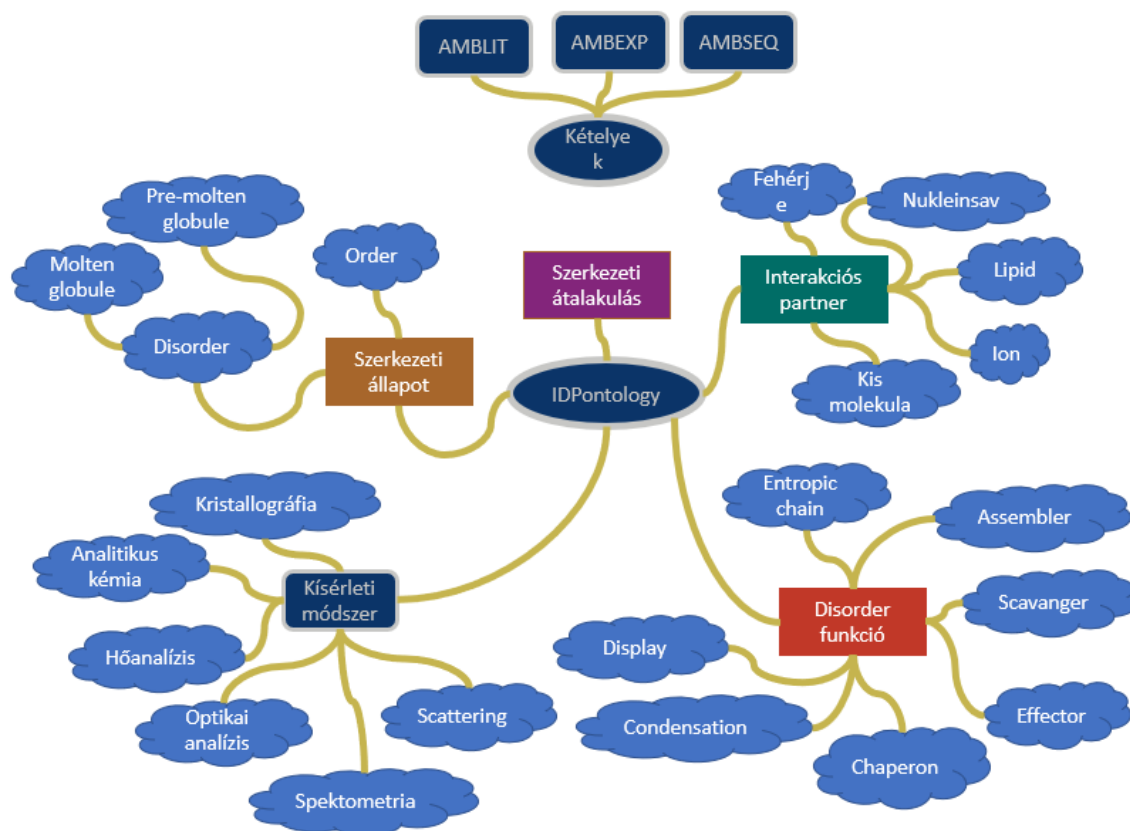
A DisProt adatbázis segítségével azonosítottam a fehérjék rendezetlen régióit. Ezek az információk kulcsfontosságúak voltak a kiindulási szerkezetek összeállításánál, mivel lehetővé tették, hogy a rendezetlen régiókat megfelelően modellezem és rendezetlenként kezeljem az integratív modellezés során.

Az ilyen rendezetlen fehérjék esetén a lényegi kérdés az, hogy ezt a rendezetlenséget valóban megtartják-e az adott sejten belül is. Az ezekre vonatkozó vizsgálatokat szerencsére ez az adatbázis tartalmazza. Ilyen és ehhez hasonló kísérleteket végeztek el többek között például a Parkinson kór esetében is mely során az α -synuclein (DisProt DP00070) fehérjének szerepe van [119]. Ez a szerep abból adódik, hogy különböző kísérletekkel igazolták, hogy nagy valószínűséggel ez a rendezetlenség azért van, hogy amikor adott fehérjéhez kapcsolódnak a rendezetlen fehérjék, akkor a komplex új funkciójához szükséges térszerkezetet fel tudják venni kötés-mediált folding során [120]. Az adatbázis közel 1.6k fehérjét tartalmaz és közel kétszer ennyi régiót. Az adatbázis össze van kötve a Pfam adatbázis domén információival és ahhoz képest adja meg a rendezetlen régiókat.

7.1.3 Pfam

A Pfam (<http://pfam.xfam.org/>) adatbázis elsődlegesen a fehérjék doménszerkezetének feltérképezésére alkalmas adatbázis [121]. A fehérje megismerésének első lépései közé tartozik, hogy milyen doméneket tartalmaz, mert ezek lesznek a főkötőhelyek.

A Pfam adatbázisban található információk segítségével pontosan azonosítottam a



7.2. ábra. **Főbb tudnivalók a DisProt elemek felépítéséről.** Ez az ábra a DisProt adatbázis elemeinek felépítését mutatja be, amely az intrinszikusan rendezetlen fehérjék (IDP) és azok tulajdonságainak kategorizálására szolgál. Központi eleme az **IDPontology**, amelyhez több különböző kategória kapcsolódik. A **Szerkezeti állapot** kategóriában különböző konformációs állapotok találhatóak, mint a **Disorder**, **Molten globule**, **Pre-molten globule** és **Order**, amelyek a fehérjék rendezettségi állapotait jelölik. A **Kísérleti módszerek** kategóriába tartozik a **Kristallográfia**, **Analitikus kémia**, **Hőanalízis**, **Optikai analízis**, **Spektrometria** és **Scattering**, amelyek a fehérjék szerkezeti és funkcionális vizsgálatára szolgáló technikák. A **Disorder funkció** kategória bemutatja a rendezetlen fehérjék különböző szerepeit, mint például a **Display**, **Condensation**, **Entropic chain**, **Assembler**, **Scavenger**, **Effector** és **Chaperon**, amelyek a fehérjék biológiai funkcióit és interakcióit írják le. Az **Interakciós partnerek** kategóriában olyan molekulák szerepelnek, amelyekkel a fehérjék kölcsönhatásba lépnek, ideértve a **Fehérjéket**, **Nukleinsavakat**, **Lipidokat**, **Ionokat** és **Kis molekulákat**. A **Szerkezeti átalakulás** kategória azokat a folyamatokat jelzi, amelyek során a fehérjék szerkezete megváltozik. Végül a **Kételyek** kategóriában szerepelnek az **AMBLIT**, **AMBEXP** és **AMBSEQ** elemek, amelyek az adatbázisban található adatok bizonytalanságait vagy vitatott pontjait jelölik. Az ábra a szerző saját munkája.

fehérjék doménjeit és kötőhelyeit. Ez lehetővé tette, hogy megértssem, mely régiók fontosak a fehérje funkciójának szempontjából, és hogyan lépnek interakcióba más fehérjékkel vagy

molekulákkal. Továbbá a Pfam adatbázis segítségével homológ szerkezeti doménelemeket kereshettem, amelyek hasonlóak a vizsgált fehérje ismert részeihez. Ez lehetővé tette, hogy a hiányzó szerkezeti információkat homológia alapján pótoljam, biztosítva ezzel a modell teljességét és pontosságát.

Több módon is lehetőségünk van doméneket keresni: például fehérjeszekvencia vagy PDB térszerkezet alapján. A keresés és összehasonlítás többszörös szekvencia igazításokon és Hidden Markov Modellen alapszik. A találatokat klánokra osztja szekvencia-, szerkezeti hasonlóságok és HMM-profilok alapján [122], de elsődlegesen a klánok osztályozása evolúciós kapcsolatok szerint történik. A klasszifikáció validálása más (pl: CATH) adatbázisok összevetésével történik. Jelenleg 628 klánt tartalmaz. A HMM-profilokat a HMMER (<http://hmmer.org/>) szoftware segítségével készítik, mely mára a BLAST sebességével vetekszik, miközben valószínűségi modellekre támaszkodik. A találatokat a Pfam hat típusra osztja: család, domén, motif, repeat, coiled coil vagy rendezetlen [122]. A doménstruktúrán kívül a szekvenciáját, lehetséges interakciókat, a fajok fáját és szerkezeteket is találhatunk a keresett doménhez. Amennyiben a domén jól definiált a szakirodalom által, akkor az ehhez kapcsolódó cikkeket és leírásokat elérhetjük egy kattintással.

7.1.4 Gene Ontology

A Gene Ontology (<http://geneontology.org/>) [123],[124] tudásbázis a világ legnagyobb információforrása a gének funkciójáról.

A Gene Ontology adatbázis segítségével azonosítottam a posztzinaptikus denzitásban (PSD) található fehérjékhez kapcsolódó géneket és ezek expressziós mintázatát, valamint az esetleges ismert mutációkat.

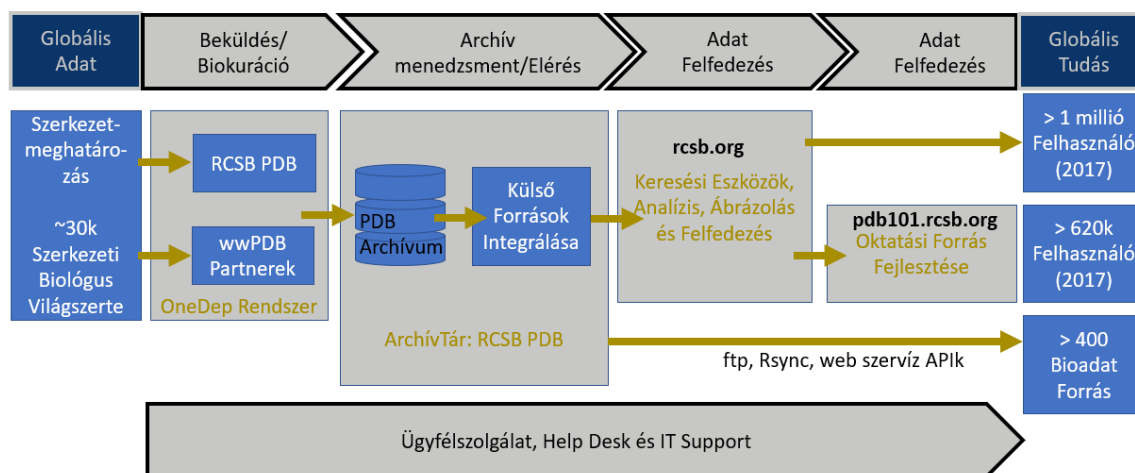
Maga az ontológia egy információmegosztási protokoll. Az ontológiát a számítástudományból importálták. Célja a biológiai adatok gyűjtését, rendszerezését és közzétételét irányító különféle fogalmi keretek leírása [125]. A fogalmak megfelelő rendszerezése elengedhetetlen, ahhoz, hogy egységes keretek között tudjunk a témában kutatni, hasonlóan az objektumorientált programozáshoz, vagy az adatbázisok felépítéséhez. Meg kell lennie a fogalmak viszonyának, szülő és gyermek fogalmaknak, hogy mely fogalom mit jelent, mikor használhatjuk, milyen tulajdonságokat hordoz magában - ezek a tulajdonságok örökölhetőek. Leegyszerűsítve sok kisebb molekuláris funkció egy adott helyen magával hozza a biológiai funkciót, mert azok alapján fejeződik ki a biológiai hatása egy egy molekulának, így az őt hordozó génnek is.

Alapja a nagyszabású molekuláris biológiai és genetikai kísérletek számítógépes elemzésének az orvosi biológiai kutatások során. A fehérjék géninformációi között találunk filogenetikai fát, valamint expressziót befolyásoló faktorokat is, biológiai folyamatokat, előfordulási helyeket, és talán számomra a legérdekesebb: olyan útvonalakat, melyek adott betegséghez tartoznak és a fehérje szerepet játszik benne. Szintén jó segítség, hogy az ortológ és paralóg géneket is megtaláljuk, így amennyiben például térszerkezet hiányzik a legközelebbi ortológoknak is utána tudunk nézni. Számos a gén funkciójával kapcsolatos cikket találhatunk felsorolva. Ezt a PANTHER-en keresztül ismerhetjük meg, melyre a GO keresés visz.

7.1.5 RCSB - Protein Data Bank (PDB)

Az RCSB Protein Data Bank (<https://www.rcsb.org/>) a fehérjék háromdimenziós térszerkezeteit annotálja és gyűjti egybe [126]. Az adatbázis jellegzetessége, hogy külön oktatási

célra is gyárt anyagokat és céljának tartja az oktatás fejlesztését. A 7.3. ábra szemlélteti az adatbázis logikai felépítését. A kutatók meghatározzák kísérletek segítségével (Röntgen kristallográfia, NMR) a fehérjék adott térszerkezetét, ezt publikálják, feltöltik. A térszerkezetek bekerülnek az adatbázisba, melyeket biokurátorok annotációkkal látnak el és igyekeznek validálni. Ezután a szerkezet hozzáférhető és vizualizálható, többféle fájlformátumban letölthető, nem csak PDB-ben. Ezekből az adatokból később oktatási anyagokat készítenek és terjesztenek. Adott UniProt Acc számú fehérjére több szerkezet is bekerülhet, ezeket könnyedén rendszerezve tudjuk lekérni és össze tudjuk hasonlítani, mely aminosavak hiányoznak, vagy esetleg mutálódnak a szerkezetben. Az összehasonlítás során domén annotációk is integrálva vannak.



7.3. ábra. **Az RCSB Protein Data Bank szerkezeti felépítése.** A folyamat a globális adatok beküldésével és biokurációjával kezdődik, ahol szerkezeti meghatározások történnek, és az adatokat a RCSB PDB és wwPDB partnerek a OneDep rendszer segítségével kezelik. Ezek az adatok a PDB archívumban kerülnek tárolásra, amely integrálja a külső forrásokat is. Az archívumból az adatokat keresési eszközök, elemzési és ábrázolási lehetőségek révén teszik elérhetővé a rcsb.org oldalon, valamint oktatási célokra is felhasználják a pdb101.rcsb.org platformon. Az adatokhoz való hozzáférést ftp, Rsync és web szervíz API-k biztosítják. Az ábra kiemeli, hogy 2017-ben több mint 1 millió felhasználó és több mint 400 bioadat forrás használta a PDB-t, jelezve annak globális jelentőségét és széles körű alkalmazását a tudományos közösségben. Az ábra a [127] cikk ábrája alapján lett magyarítva.

7.1.6 Cath

A CATH (<https://www.cathdb.info/> leginkább doménszerkezetek összehasonlítására és klaszterezésére alkalmas [128]. Általa megtudhatjuk, hogy az adott domén, amivel foglalkozunk, mennyire különbözik a hozzá legközelebb álló, azaz a vele egy családban levő doménektől. Nagy különbségek esetén például nehezebben "cserélhetők" ki a domének és valószínűleg a funkciójuk sem lesz olyan közeli. Adott doméncsaládra jól összefoglalja különböző adatbázisok terminológiáit és kördiagramról könnyen leolvasható, hogy a család tagjai milyen arányban kaptak adott annotációt, így például GO adatokat is leszűrhetünk a családról [129].

A CATH rendszere a következő felosztást követi: Class-Architecture-Topology-Homológ család.

7.1.7 BrainSpan Atlas

BrainSpan Atlas of the Developing Human Brain [130]. Az Allan Brain Atlaszok igyekeznek feltérképezni mind az emberi, mind az egér agynak különböző régióit és a bennük lévő gének expressziós szintjeit. Ezáltal kísérletes adatokat tárolnak agyi régiókra és alanyokra lebontva. Az atlaszok nem hibátlanok és nagyon függenek az elvégzett kísérletektől, viszont jó alapot szolgáltatnak arra, hogy modellemben használhassam őket. Szeretném kiemelni, hogy mRNS expressziós szinteket tartalmaznak, amelyeket még át kell konvertálni fehérjeabundanciákra.

7.1.8 Postsynaptic Interaction Database - PSINDB

Az adatbázist Kálmán Zsófia fejlesztette ki karunkon [131], ami már ténylegesen a posztzinaptikus denzitás fehérjéire koncentrálna gyűjti össze a szükséges interakciókat. Az adatbázis összegyűjti az ismert interakciókat és még meg is mutatja a kötési régiókat, melyeket fel tudok használni akár mutációanalízisnél, akár a háromdimenziós szerkezetek modellezésénél. Az interakciók mellett kigyűjti a fehérjéhez társított funkciókat is. Ami szintén előnye az adatbázisnak, hogy a fehérjék adott mutációit tartalmazó blokkal rendelkezik. Ebben megismerhető, mely aminosavak mely aminosavakra cserélődnek, és mely fenotípus megjelenéséhez lehet közük. Azonban fehérjestabilitás megváltozásokat, ezáltal kötési erősségek megváltozását az adatbázis már nem tartalmazza. Épp ugyanígy az expressziós szintek és mértékük nincsen az adatbázisban, annak bevétele mégjobban segítheti a munkámat. Ez mindazonáltal előzetes vizsgálataim szerint nehézkes, hiszen számos különböző kísérlet van, össze nem egyeztethetőek ezáltal a különböző abundancia adatok stb. Az adatbázis hátránya az azonban, hogy még nincsen minden adat feltöltve és hiányos. Hosszútávon viszont ideális lesz a kutatás egészéhez.

7.2 Fehérje- és kötéstabilitást jósló módszerek

A fehérjék kötési erőssége és stabilitása kulcsfontosságú tényező a biológiai folyamatok megértésében és gyógyszerfejlesztésben. Az előrejelzések lehetővé teszik számunkra, hogy megértsük, hogyan működnek, kapcsolódnak össze ezek a makromolekulák és ezáltal a posztzinaptikus denzitás hogyan változik különböző körülmények között, ami hozzájárulhat a betegségek mechanizmusainak feltárásához és hatékonyabb terápiák kifejlesztéséhez. Hiszen, ha nem vesszük figyelembe a kötési erősségek különbözőségét, akkor minden interakció ugyanolyan valószínűséggel következne be, amiről tudjuk, hogy nem így van.

Ha nincs rendelkezésre álló kísérleti adat, akkor először számítógépes módszerekkel kell becsléseket készítenünk. Ezeket a becsléseket később a kutatócsoport kísérletesen fogja validálni. Ilyen eszközök lehetnek, amelyeket megfontoltam például a HoTMuSiC [132] ami előrejelzi a pontmutációk termálstabilitását amennyiben a funkcionalitás változatlan marad. A módszer elkerüli a jól konzervált vagy túlságosan doménközeli tartományokat, mert az előbbi bekövetkeztének kicsi a valószínűsége, míg az utóbbi esetén nagy a valószínűsége, hogy a domén tulajdonságai megváltoznak, így a fehérje funkciója nem maradna ugyanaz. Ez azt jelenti, hogy a HoTMuSiC alkalmas olyan mutációk kiszűrésére, amelyek esetünkben nem befolyásolják a kötési sebességet, tehát nem szükséges vizsgálnunk.

Az online PredictSNP [133] szerver több előrejelzési módszert egyesít, és szüksége van egy aminosavszekvenciára FASTA fájlformátumban, és manuálisan megadott mutációs adatokra, ha ezen mutációs adatok nem állnak rendelkezésre. A lehetséges mutációk megtalálásához a BioMuta [134] adatbázis elsősorban a rákkal kapcsolatos egypontos mutációkhoz érhető el. Az online PredictSNP eszköz egyik hátránya, hogy sok (több mint 100) munka van előtte egy állás beküldésekor – eltarthat egy ideig, mire megérkezik az eredmény. Az eredmény a mutációk magabiztosságát és patogenitását adja. Például lefuttattam egy tesztet PSD-95-re egy R65K mutációval, amely normál patogenitást eredményezett 74%-os megbízhatósággal, ami azt jelenti, hogy nem befolyásolja a funkciókat és a stabilitást. A többi módszer csak összehasonlításra szolgál, és a legtöbbnek hasonló eredményt kell adnia az elfogadáshoz.

7.2.1 CAD-pontszám

A CAD-pontszám (Contact Area Difference, CAD-score) [135] háromdimenziós struktúrák hasonlóságát leíró valószínűségi mérték, amely az egyes struktúrákon belüli belső érintkezési felületek megváltozása alapján számítható ki. Azért szeretném kiemelni, mert a modellezéseink során is kiszámítható és használható mérték. A hasznos pont az, hogy nulla és egy közötti normalizált értéket ad, amely megmondja, hogy a modell mennyire van elég közel a cél-referenciaszerkezethez. 1 esetén a szerkezeten belüli kontaktok teljesen megegyeznek, míg a 0 esetén teljesen különböznek, vagy akár semmilyen érintkezés sincsen a szerkezeten belüli aminosavak között. Amennyiben a vizsgált tartományt leszűkítjük csak a kötési régióra, akkor ez az érték megadhat egy számot, amely jelzi, hogy milyen mértékben változott meg a kötési erősség.

A CAD-scoret az alábbi módon kell kiszámolni [136]:

$$\text{CAD-score} = 1 - \frac{\min \sum_{(i,j) \in G^*} \{|T_{(i,j)} - M_{(i,j)}|, T_{(i,j)}\}}{\sum_{(i,j) \in G^*} T_{(i,j)}} \quad (7.1)$$

A target, célszerkezetben szereplő i - j aminosav közti kontaktfelület a $T_{(i,j)}$, a modelben szereplő kontaktfelület az $M_{(i,j)}$, a $G^* \subset G$ a kötési régióban szereplő aminosavak descart szorzata. (Az eredeti képletben G szerepel, azaz a szerkezet összes aminosavpárját tekinteték.) A minimumkeresésre a képletben azért van szükség, mert abban az esetben, ha a modell kontaktfelülete meghaladná a kétszeresét a target kontaktfelületének ($M_{(i,j)} > 2T_{(i,j)}$), akkor nem valószínűségi mértéket kapnánk. A kontaktfelületet érintkező atomok segítségével számolják ki: Két atom érintkezőnek tekinthető, ha a Voronoi tesszellációban szomszédosak, és a vízmolekula ($r = 1.4\text{\AA}$) nem fér el közöttük [136].

A G^* módosítással kötési régió megváltozására kapunk egy valószínűségi mértéket. Ebben az esetben az adott referenciaszerkezet lehet egy komplex, melyben két vad típusú fehérje összekapcsolódik, míg a modell, amelyet hasonlítani szeretnénk a referenciához lehet ugyanez a komplex azzal a különbséggel, hogy az egyik fehérje szerkezetében mutációt idézünk elő.

A predikció hátránya, hogy szükségünk van vad típusú és mutáns típusú kötések tartalmazó szerkezetekre – csak a domén-domén struktúra lehet alkalmas, mivel a fehérje egy részhalmazán dolgozunk. Ha rendelkezünk vad típusú és mutáns fehérjék szerkezetével, akkor mind a vad típusú, mind a mutáns fehérjék kötődését modellezhetjük IMP segítségével, majd kiszámíthatjuk a modellezett kötések CAD-pontszámát, amellyel megállapítható, hogy a kötést mennyire változtatta meg a mutáció. A módszer korlátjai az IMP modellezési

hibában, a CAD-score hibában és a CAD-pontszámból lefordított kötési erősségben rejlik, tehát a korlátokat a kiértékelésnél figyelembe kell venni. Szintén figyelni kell, mert a számítás a megváltozás mértékére ad egy predikciót, de annak irányára önmagában nem.

7.2.2 NeEMO

Azért esett a választásom a NeEMO online is elérhető eszközre, mert számításai során számos megfigyelést integrál egybe, ami megegyezik kutatásom lényegével. A NeEMO [137] stabilitás megváltozását értékeli úgy, hogy a fehérjében létrejövő kölcsönhatásokat kölcsönhatási hálózatok segítségével modellezi. Az aminosavak energiahozzájárulását többek között a torziós szögek kiszámításával számítja ki. Másodlagos szerkezeteket DSSP-vel predikál. Adott PH-n és hőmérsékleten dolgozik. A térszerkezetek figyelembevételével mellett többszörös szekvencia illesztésekkel konzervált régiókat határoz meg. Ezen különböző adatokat átkonvertálva többdimenziós vektortérre betölti egy neurális hálózatba, melynek kimeneteként a stabilitás változását adja meg Kcal/mol-ban. A stabilitásváltozás a vad típusú és a mutáns típusú Gibbs-szabad energia különbsége, így kaphatunk irányt és nagyságot is.

$$\Delta\Delta G = \Delta G_{\text{WT}} - \Delta G_{\text{MT}} = -RT \ln \left(\frac{[\textit{denaturated}]}{[\textit{naturated}]_{\text{WT}}} \right) + RT \ln \left(\frac{[\textit{denaturated}]}{[\textit{naturated}]_{\text{MT}}} \right) \quad (7.2)$$

Ha a $G > 0,5$, akkor a mutáció destabilizáló hatású.

Egy leegyszerűsített példán szemléltetve jobban megérthetőek az értékek jelentései. Vegyünk 1000 darab házat, amelyet felépítettünk. Ebből omoljon össze 4. Majd végezzünk el egy módosítást a tervrajzon. A módosítás után újabb 1000 ház felhúzása után azt tapasztaljuk, hogy 50 dőlt össze. A tervrajzon való módosítás valószínűsíthetően destabilizálta az épület szerkezetét. Ezt a számokkal nézve:

$$\Delta\Delta G = -\ln \left(\frac{4}{996} \right) + \ln \left(\frac{50}{950} \right) = 2.57 \quad (7.3)$$

A példában a gázállandó és a hőmérséklet általi arányossági tényezőtől eltekintettem, mert nem molekulát néztem.

Ha $0,5 > G > -0,5$ intervallumba esik, akkor hibahatáron belül van az érték, így nem meghatározható, hogy valóban megváltozott-e a stabilitás - változatlanak tekintjük. Példánkban ez azt jelenthetné, hogy, ha eredetileg 4 épület dőlt össze, a módosított tervrajzúakból pedig 6 épület dőlt össze, akkor az érték a következőképpen alakulna:

$$\Delta\Delta G = -\ln \left(\frac{4}{996} \right) + \ln \left(\frac{6}{994} \right) = 0.4 \quad (7.4)$$

Ha $G < -0,5$, akkor a mutáció stabilizálja a fehérjét. Például a módosított épületek közül csak 1 dőlne össze.

$$\Delta\Delta G = -\ln \left(\frac{4}{996} \right) + \ln \left(\frac{1}{999} \right) = -1.4 \quad (7.5)$$

A szervert használom arra, hogy a mutációk kötése erősségre mért hatását becsüljem addig, amíg a kutatócsoportunk nem biztosít kísérletes adatokat számunkra.

7.3 A Gillespie algoritmus, amelyet a Cytocast szoftver is használ

A fehérjekomplex képződés szimulációjához először a SiCompre, majd a továbbfejlesztett Cytocast (cytocast.com) ágens alapú szimulációs eszközt használtam. A Cytocast modellezési megközelítést kínálja a szinaptikus denzitás diverzitásának vizsgálatára fehérje komplexek komponenseik abundanciáját tekintve. A Kheömley-Horváth Bence Márk által kifejlesztett Cytocast szoftver, amelyet használok a Gillespie-algoritmus egy változatát valósítja meg. Ezért a lépések száma nem lineárisan skálázódik a szimuláció futási idejével. Az algoritmusnak ki kell számítani a saját futási idejét, azaz a következő reakcióig eltelt időt. Ez egy Monte Carlo alapú módszer minden reakció szimulálására, hogy a molekulák vagy esetünkben komplexek mennyiségi változásának nem determinisztikus megközelítésével élhessenek. A Cytocastra azért van szükségem, mert millió komplexképződés lehetséges figyelembe véve a számos kötési lehetőséget, ezáltal gyakorlatilag lehetetlen minden egyes reakciót determinisztikus differenciálegyenletekkel felírni.

A Gillespie-algoritmus [138] lépései:

1. Állítsa be a kezdeti feltételeket a kiindulási molekulák számával (itt a fehérje bősége) $n_{i,0}$ és a lehetséges reakciószámokkal q
2. Két valószínűségi változó generálása 0 és 1 között. (r_1, r_2)
3. Számítsa ki az egyes reakciók hajlamfüggvényeit (itt kötés, jegyezzük meg, ha kötődés történik, eggyel kevesebb fehérje van).

$$\alpha_i(t) = (n_{i,t-1} - 1)k_i \quad (7.6)$$

4. Számítsa ki a hajlamfüggvényt a teljes rendszerre:

$$\alpha_0 = \sum_{j=1}^q \alpha_j(t) \quad (7.7)$$

5. Számítsa ki az időt, amikor a következő kémiai reakció lezajlik: $t + \tau$ ahol

$$\tau = \frac{1}{\alpha_0} \ln \left(\frac{1}{r_1} \right) \quad (7.8)$$

6. Számítsa ki, hogy melyik reakció játszódik le, majd állítsa be a fehérjék számát (egy fehérjével kevesebb a bemenetről és eggyel több a kimenetről (komplex)). A i^{th} reakció akkor következik be, ha a feltételek igazak:

$$\frac{1}{\alpha_0} \sum_{j=0}^{i-1} \alpha_j \leq r_2 < \frac{1}{\alpha_0} \sum_{j=0}^i \alpha_j \quad (7.9)$$

A Cytocast futtatásához a bemeneti adatokat egy-egy kísérleti, nagy mennyiségű adathalmazból kell kinyerni és a megfelelő formátumra hozni. A Cytocast az elődjénél hatékonyabban képes a sok különböző bemeneti adatot tartalmazó szimulációkat lefuttatni. Bár a SiComPre is indítható terminálból is, mégis a SiComPre esetében általában manuálisan kellett megadni a különböző összetett bemeneti fájlokat, míg a Cytocast esetében jobban lehetőségem nyílt arra, hogy egységes bemeneti fájlokat generáljak, majd ezek köré írtam egy Python kódot, ami segítségével tudtam kezelni a több száz szimulációt, annak lefuttatásait és azok eredményeinek megfelelő rendezését.

A Cytocast használata összességében több előnyt nyújt, mint a SiComPre. Kimutatták, hogy a Cytocast képes hatékonyan modellezni a fehérje komplex eloszlását akár teljes

sejtekben is.

SiComPre [139] Előnyök:

- Gillespie-algoritmus jól definiált
- Kvantitatív szimuláció
- Az ágensek sztochasztikus diffúziója jól definiált
- Meglevő, kész, stabil szoftware
- Könnyen kattintható felület

Hátrányok:

- Nagymennyiségű szimuláció során kedvezőtlen
- A domének 3D orientációjából következő restrikciónak (hozzáférhetőség) nem adhatók meg

Cytocast Előnyök:

- Gillespie-algoritmus jól definiált
- Kvantitatív szimuláció
- Háromdimenziós térben mozgó ágensek sztochasztikus szimulációja
- Új lehetőségek implementálhatóak benne, mert fejlesztés alatt áll:
 - Membrán
 - Dinamikus domének
 - Kisebb molekulák pl ionok, amik befolyásolhatják a komplexeket, hozzáadhatóak lehetnek majd a jövőben.

Hátrányok:

- A domének háromdimenziós orientációja nincs implementálva benne, mert pontként kezelték a fehérjék.

7.4 K-means algoritmus

A K-közép klaszterezés az egyik legegyszerűbb algoritmus hasonlósági osztályok, úgynevezett klaszterek létrehozására egy adott térbeli közelség alapján. Olyan algoritmusokkal, melyek a régiók hasonlóságát vizsgálják szimulációk kereshetők és elemezhetők bemeneti adatok és eredmények alapján is.

A k-means algoritmust gyakran használják az adatbányászatban és gépi tanulásban. Az algoritmus célja, hogy minimalizálja az egyes klasztereken belüli varianciát [140].

1. **Klaszterközéppontok inicializálása:** Válasszunk ki véletlenszerűen k adatpontot (régión), amelyek az első klaszterközéppontok (centroidok) lesznek.
2. **Adatpontok hozzárendelése:** Minden adatpontot, azaz régiót rendelünk ahhoz a klaszterhez, amelynek középpontja a legközelebb van az adatponthoz. A közelséget mind a bemeneti adatok esetében, mind a kimeneti adatok esetében euklideszi távolság alapján mértem, az előbbiben a fehérjék abundanciája, míg az utóbbiban a létrejövő komplexek abundanciája alapján alkotott vektorterekben dolgoztam.
3. **Klaszterközéppontok újraszámítása:** Az új klaszterközéppontokat úgy számoljuk ki, hogy meghatározzuk az egyes klaszterekhez rendelt adatpontok átlagát.
4. **Konvergencia ellenőrzése:** Ellenőrizzük, hogy a klaszterközéppontok pozíciói megváltoztak-e az előző iterációhoz képest. Ha nem, az algoritmus konvergált és leáll. Ha igen, térjünk vissza a 2. lépéshez.

Az algoritmust a class sklearn.cluster.KMeans osztállyal valósítottam meg Pythonban.

Túl sok klaszter használatával még a kis különbségek is nagy változásnak tűnhetnek. A K-means klaszterezés [141] során minél nagyobb a csoportszám, annál több hibalehetőség merülhet fel, hasonlóan a túlillesztés azaz angolul az overfitting jelenséghez. Minél nagyobb a keresett csoportszám, annál bizonytalanabbá válik a határon lévő pontok, régiók helyzete. Ebből adódóan már minimális elmozdulás is jelentős változásnak hathat rajta. Ez félrevezető

eredményt szülhet, mert a sztochasztikus zajok kiszűrése sosem maximális - túl sok csoport esetén pedig ez a megmaradt zaj erősödik fel. Túl kevés csoportszám esetén ennek az ellenkezőjét tapasztalhatjuk, azaz csekély információtartalommal fog bírni, hogy melyik régió melyik csoportba kerül, mert a túl távoli szimulációk is közös csoportba fognak kerülni [142]. Tehát a sztochasztikus szimulációk zajának csökkentése érdekében kevesebb klasztert kell használni annak ellenőrzésére, hogy vannak-e szignifikáns változások, de így a kisebb változtatások nagyobb jelentőséggel bírhatnak.

Ezért K-means eljárás során éppúgy, mintha egy neurális hálózatot szeretnénk a problémánkra illeszteni - ami a későbbiekben célom is - oda kell figyelni és gondosan kell kiválasztani, hány csoporttal szeretnénk dolgozni. Tehát a K az egyfajta metaparaméter, melyet szintén az adatokhoz kell illesztenünk.

A K-means klaszterezés során, hogy elkerüljem a már említett overfitting jelenséget, ebben a tanulmányban K-közép klaszterezést végeztem K=2,3,4,5 és 6 esetén (lásd a kiegészítő anyagot), így kiszűrve a maradék sztochasztikus zajt. Mindazonáltal az, hogy nem annyi klasztert készítettem, amennyi régiótípus van azzal járt együtt, hogy a K-means klaszterezés nem vesz figyelembe direkt összehasonlítható jellegzetességeket - természetesen indirekt figyelembe veheti őket - azaz sem agyi régió típusonként, (például elsődleges vizuális cortex), sem pedig alanyonként nem tudom direkt csoportosítani általa a jelenségeket - bár ábrázolásban még megfigyelhetem őket, melyik régió melyik csoportba kerül.

Az ideális klaszterszámot a silhouette score segítségével határozhatjuk meg. Egy adott szimuláció c_i sziluett együttható értéke [143]:

$$\frac{b_i - a_i}{\max(b_i, a_i)} \quad (7.10)$$

ahol az a_i az átlagos távolság az adott szimuláció v_i és az azonos klaszterben C_I levő szimulációk között,

$$a_i = \frac{1}{|C_I| - 1} \sum_{v_i, v_j \in C_I, i \neq j} |v_i - v_j| \quad (7.11)$$

b_i pedig a v_i legkisebb átlagos távolsága a többi klaszter pontjától, azaz minden másik klaszter esetén átlagoljuk az adott szimulációs távolságokat klaszterenként, majd a legkisebbet választjuk. Gyakorlatilag a "legközelebbi" klaszter pontjaitól való átlagos távolság.

$$b_i = \min \left\{ \frac{1}{|C_J|} \sum_{v_i \in C_I, v_j \in C_J, i \neq j} |v_i - v_j| \right\} \quad (7.12)$$

- A +1-hez közeli értékű sziluett pontszám azt jelenti, hogy az adatpont a megfelelő klaszterben van.
- A 0-hoz közeli értékű sziluett pontszám azt jelenti, hogy az adatpont egy másik fürthöz tartozhat.
- A -1-hez közeli értékű sziluett pontszám azt jelenti, hogy az adatpont az rossz klaszterben van.

Tehát a cél, hogy minél több szimuláció sziluette értéke legyen minél magasabb. Ez általánosan bármely más típusú klaszterezés után alkalmazható annak milyenségét ellenőrizendő.

7.5 Főkomponens Analízis

Főtengelytranszformáció, más néven a Főkomponens Analízis [144] (PCA - principal component analysis) egy statisztikai módszer, amelyet az adatkészlet dimenziójának csökkentésére használnak, miközben a lehető legtöbb eltérést megőrzik az adatokban. Ez egyfajta lineáris transzformáció, amely korrelált változók halmazát veszi fel, és nem korrelált változók halmazává (a fő komponensekké) alakítja át. Ezt a technikát az adatok mintázatainak azonosítására, a zaj csökkentésére és az adatok könnyebben láthatóvá tételére használják. A PCA bármilyen típusú adatra alkalmazható, beleértve a képeket, a hangot, a szöveget és a numerikus adatokat.

Főtengelytranszformáció, során direkt csoportosítás nem történik, viszont sokkal látványosabb ábrákat tudok kapni azáltal, hogy a legnagyobb információtartalmú tengelyek által kifeszített síkon jelenítem meg a szimulációkat. Az adathalmazom esetében egy olyan vektortér, mely vagy a bemeneti abundanciák által vagy a kimeneti komplexek abundanciája alapján elhelyezi az összes régiót. Az adathalmaz vektorterén bázistranszformáció történik. Az új bázis a sajátvektorok bázisa sajátértékeik szerinti sorrendben. Ekkor a legelső két bázisvektor kifeszíti a leginformatívabb síkot, melyen ábrázolhatom a régiókat.

Végső soron csoportosulások megjelenhetnek rajta, anélkül, hogy mesterségesen próbálnánk őket csoportosítani. Pontosabb eredmények eléréséhez akár az így kapott és transzformált pontokat is lehet K-means klaszterezni és megnézni, így milyen eredményeket kapunk. A főtengelytranszformáció a sokdimenziós adatteret levetíti a leginformatívabb síkra, ezért alkalmas a kiugró adatok észrevételére, mind a bemenet kísérletes adatok tekintetében, mind pedig a kimeneti adatok vad típusú, vagy mutáns vizsgálata során.

Minden PCA a `scikit.learn.decomposition.PCA` segítségével történt Pythonban. A függvény `single value decomposition-t` (SVD) használ.

$$M = U\Sigma V^T$$

Az M valós mátrix tartalmazza mind az 524 régió oszlopvektorait. $n \times n$ -es mátrix esetén nagyon egyszerűen a $n \times n$ -es mátrixokra bomlana fel, ahol a középső mátrix tartalmazza a sajátértékeket, a szélső mátrixok a sajátvektorokat - valóban egy leképezés bázisvektora lenne. Esetünkben a bemeneti adatokat tekintve a mátrix 524×7 -es alakú. Ekkor értelemszerűen a szorzatnak is annak kell lennie. A Σ mátrix egy nem négyzetes mátrix, de továbbra is diagonális, melynek a főátlójában az önadjungált operátor (M^*M) (feltétlenül nem negatív) sajátértékeinek négyzetgyökei vannak csökkenő sorrendben - tehát az analógia továbbra is él. Ezen értékek mondják meg, hogy az adott sajátvektor (nem négyzetes esetben itt singular vektor) irányában a leképezés mennyire nyújt - tehát mennyire fontos a tengely.

A két legfontosabb főtengely (singular vektor) bázisára áttérve egy síkon ábrázolhatóvá válik minden régió, melyben egymáshoz képesti viszonyukat összevethetjük. Figyeljünk arra, hogy külön lesz egy ilyen transzformáció a bemenetekre és a kimenetekre - két különböző vektortérben dolgozom, amely miatt az összevetéshez további megfontolásokra van szükség.

7.6 t-próba

A t -próba leggyakoribb alkalmazása annak tesztelése, hogy két populáció átlaga különbözik-e. A kutatásom során ez a populáció például adott komplex darabszáma adott szimuláció

során. A kérdés, hogy például a mutáns fehérje esetén a 40-szer megismételt szimuláció eredményeként kijött adott komplex átlagos darabszáma ténylegesen különbözik-e a vad típusban megfigyelt átlagos darabszámtól.

Azért nem fogadható el t -próba nélkül csupán a nyers átlagokról beszélni, mert a szimuláció sztochasztikus, zajos, a rendszer ilyen értelemben nem ideális, csak közelíti a differenciál-egyenletrendszerek megoldásait. Lehet, hogy két különböző átlagot kapunk, de valójában ezt a különbséget "csak a zaj okozza" és ideális esetben, ha ki lehetne szimulációk nélkül is számolni, ugyanazt a darabszámot kaptuk volna.

A t statisztika annak tesztelésére, hogy az átlagok különböznek-e, a következőképpen számítható ki:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_{X_1}^2 + s_{X_2}^2}{n}}}$$

amennyiben azt feltételezzük, hogy a szórás megegyezik - ez megtehető, mert megfigyeléseink szerint a szórás 1 körül mozog 40 ismétlés esetén. A számlálóban az egyes átlagok különbsége található, míg a nevezőben a gyök alatt a varianciák összege elosztva az ismétlések számával.

A `scipy.stats.ttest_rel` metódust használtam arra, hogy kiszámoljam a p -értékeket, ahol a t -score az alábbi módon számítható [145]:

$$\frac{\text{mean}(a - b)}{\sigma} \tag{7.13}$$

ahol az a és b az adott komplex abundanciáinak 40 hosszúságú tömbje vad típusú és mutáns szimulációk esetén, és a σ a standard hiba.

A p -értéket a t -pontoszámból számítják ki az alternatív hipotézistípus alapján, amely kétoldalú:

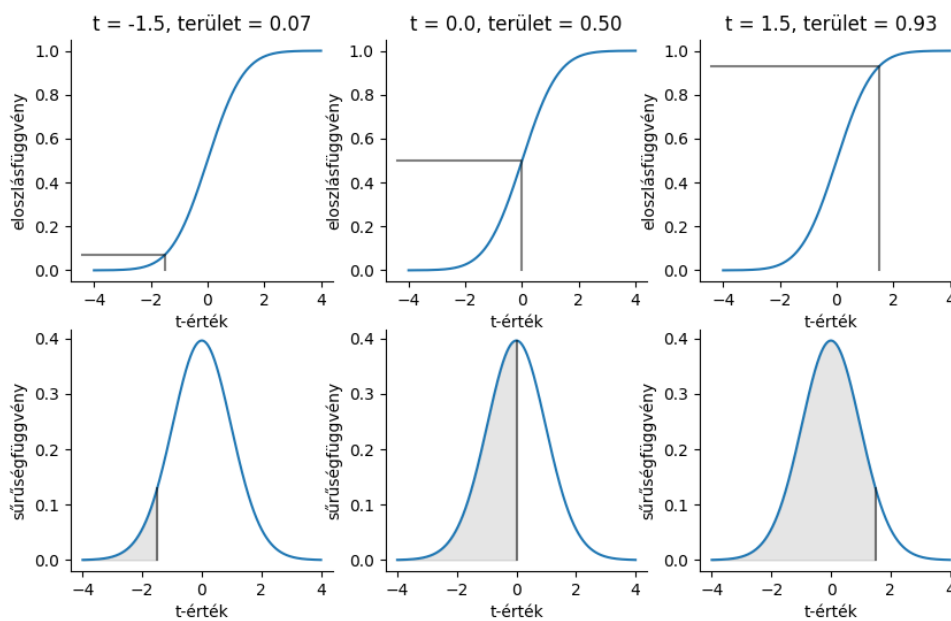
$$2cdf_{t,39}(-|t_{\text{score}}|) \tag{7.14}$$

ahol cdf a kumulatív sűrűségfüggvény a t -próbához 39-es szabadságfokkal (40 ismétlődő szimuláció esetén) lásd 7.4 ábra.

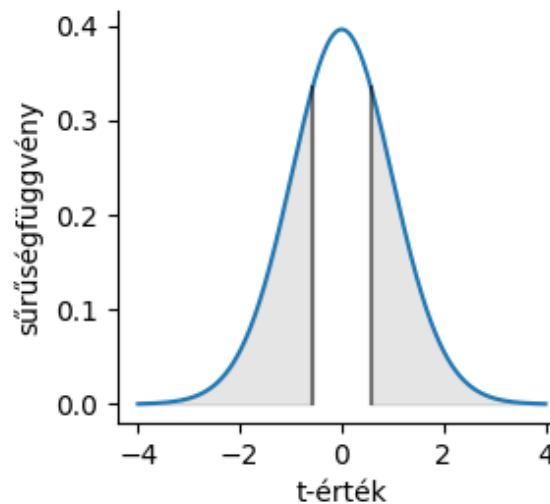
A p érték lesz a tényleges mutató, amely megmutatja, hogy van-e különbség a két adathalmaz átlaga között, vagy nem adott szignifikancia szint szerint. Fontosnak tartom azért itt kiemelni, hogy valójában a p érték csak azt tudja megmondani, hogy a nullhipotézis támogatott-e vagy sem. Nem tudja megmondani, hogy az alternatív hipotézis igaz-e, és miért. Tehát a valóságban arra kaphatunk állításokat, hogy a két átlag megegyezik elég nagy valószínűséggel, amennyiben a p érték elegendően nagy.

Ha ez a p -érték nem kisebb, mint a választott 0,05-ös szignifikancia szint, nem utasíthatjuk el a nullhipotézist. Így nincs elegendő bizonyítékunk arra, hogy kijelentsük, hogy ennek a mutáns komplexnek az átlagos darabszáma eltér a vad-típusétól.

Példával szemléltetve: A H376.IIA.50_AMY régióban az AMPAR/PSD-95/SynGAP komplex átlagok: vad típus 108.1 míg mutáns 108.9. Szórások: 6.6 és 5,9.



7.4. ábra. A 39-es szabadsági fog mellett a t -teszt eloszlás- és sűrűségfüggvénye szemléltetve: A függvény segítségével megtalálható az adott p -érték, amely a valószínűségét adja meg, annak, hogy a t érték adott érték alá esik. Éppen emiatt kétoldali próbánál ez a valószínűség kétszereződik. Az ábra a szerző saját munkája.



7.5. ábra. Kétoldalú t -teszt esetén a valószínűség a példában szereplő értékkel z X-tengelyen a t -értékek szerepelnek, míg az Y-tengelyen a sűrűségfüggvény látható. A grafikon egy normális eloszlást ábrázol, ahol a két függőleges vonal jelzi a t -értékek azon tartományát, amely alapján döntést hozunk. A besatírozott rész a p -érték. Az ábra a szerző saját munkája.

$$t = \frac{108.1 - 108.9}{\sqrt{\frac{6.6^2 + 5.9^2}{40}}} = -0.57$$

Ennek abszolút értékét kikeresendő a 39-es szabadsági foknál a p -érték 0,571947. Az eredmény nem szignifikáns, ha $p < 0,05$ (7.5 ábra).

A páros t -próba alkalmazásának kritériumai szigorú feltételekhez kötöttek, hogy a statisztikai elemzés megbízható és érvényes legyen. Először is, a párosított mintáknak két kapcsolódó csoportból kell származniuk, például ugyanazon egyedek két különböző időpontban mért eredményeiből. Másodsor, a párosított különbségek eloszlásának normálisnak kell lennie, amit gyakran a Shapiro-Wilk vagy a Kolmogorov-Smirnov tesztekkel ellenőriznek. Harmadszor, elengedhetetlen, hogy a minták véletlenszerűen legyenek kiválasztva és függetlenek legyenek egymástól, hogy elkerüljük az eredmények torzulását. Végül, a minta mérete elegendően nagy kell legyen ahhoz, hogy a statisztikai próba ereje biztosított legyen; általában legalább 10-15 páros minta szükséges. Ezen kritériumok teljesülése biztosítja, hogy a páros t -próba eredményei megbízhatóak és validak legyenek.

7.7 A Shapiro-Wilk Teszt Alkalmazása és Jelentősége

A statisztikai elemzések során gyakran találkozunk azzal a követelménnyel, hogy a vizsgált adatok normál eloszlást kövessenek. A Shapiro-Wilk teszt az egyik legismertebb és legszélesebb körben használt módszer a normalitás ellenőrzésére. Ezt a tesztet először 1965-ben fejlesztették ki Shapiro és Wilk, és azóta számos kutatás és alkalmazás során bizonyította hasznosságát és megbízhatóságát shapiro1965analysis.

7.7.1 A Shapiro-Wilk Teszt Alapjai

A Shapiro-Wilk teszt lényege, hogy a minta adatait összehasonlítja egy normál eloszlásból származó hipotetikus mintával. Az eljárás során a teszt statisztikai értékét (W) számítják ki, amely azt méri, hogy mennyire tér el a minta eloszlása a normál eloszlástól. Minél közelebb van ez az érték 1-hez, annál valószínűbb, hogy az adatok normális eloszlást követnek.

7.7.2 Alkalmazási Területek és Előnyök

A Shapiro-Wilk teszt számos előnnyel rendelkezik, amelyek hozzájárulnak széleskörű alkalmazásához a statisztikai elemzésekben. Először is, a teszt nagyon érzékeny kis mintaméreteknél is, ami különösen hasznos lehet korlátozott adatkészletek esetén. Másodsorban, összehasonlítva más normalitás tesztekkel, mint például a Kolmogorov-Smirnov teszt vagy az Anderson-Darling teszt, a Shapiro-Wilk teszt általában nagyobb statisztikai erővel rendelkezik, így jobban észleli a normál eloszlástól való eltéréseket.

7.7.3 Kritikus Értékelés és Limitációk

Bár a Shapiro-Wilk teszt számos előnnyel rendelkezik, fontos megjegyezni bizonyos korlátait is. Például nagyobb mintaméreteknél a teszt túlérzékeny lehet a normalitástól való kisebb eltérésekre, ami azt eredményezheti, hogy a teszt statisztikailag szignifikáns eltéréseket mutat olyan esetekben is, amikor ezek az eltérések gyakorlati szempontból elhanyagolhatóak. Emellett a teszt nem alkalmazható olyan adatokra, amelyeknek nullával egyenlő vagy negatív értékeik vannak, mivel a log-transzformációt igénylő adatok nem felelnek meg a Shapiro-Wilk teszt követelményeinek.

Összességében a Shapiro-Wilk teszt kiváló eszköz a normalitás ellenőrzésére, különösen kis minták esetén. Előnyei közé tartozik a magas statisztikai erő és az egyszerű alkalmazhatóság, azonban figyelembe kell venni a teszt korlátait is nagyobb mintaméretű és speciális adatkészletek esetén. A statisztikai elemzések során a Shapiro-Wilk teszt alkalmazása biztosítja a megfelelő normalitás ellenőrzését, amely alapvető fontosságú a pontos és megbízható eredmények eléréséhez.

7.8 A Szignifikancia Szint Korrigálása: Bonferroni és Benjamini-Hochberg Módszerek

A szignifikancia szint korrigálása kritikus fontosságú, amikor több összehasonlítást vagy tesztet végzünk egy adathalmazban. Ennek célja a hamis pozitív eredmények, vagyis az elsőfajú hibák (-hibák) minimalizálása. Ha több tesztet végzünk, nő annak az esélye, hogy legalább egy hamis pozitív eredményt kapunk, ezért szükségesek a korrigálási módszerek [146].

7.8.1 Bonferroni korrekció

A Bonferroni-korrekció az egyik legegyszerűbb és legismertebb módszer a szignifikancia szint korrekciójára. Ez a módszer az α szintet osztja el a végzett tesztek számával, így minden egyes teszt szigorúbb szignifikancia küszöböt kap (α/n). Bár ez hatékonyan csökkenti a hamis pozitív eredmények számát, nagy áteresztésű vizsgálatok esetén problémát okozhat. A Bonferroni-korrekció ugyanis túl szigorúvá válhat, és emiatt sok valós pozitív eredményt

is elutasíthat, csökkentve a statisztikai teljesítményt (statistical power). Ez a módszer tehát kevésbé ideális, ha nagyszámú tesztet végzünk, mivel sok potenciálisan fontos eredményt is kidobhat [147].

7.8.2 Benjamini-Hochberg korrekció

A Benjamini-Hochberg (BH) módszer dinamikusabb és adaptívabb alternatívát kínál a Bonferroni-korrekcióval szemben. Ez a módszer a hamis felfedezési arányt (False Discovery Rate, FDR) kontrollálja, nem pedig az egyes tesztek szignifikancia szintjét. A BH módszer lényege, hogy a p -értékeket sorba rendezi, majd egy lineáris regressziót alkalmaz ezekre a sorba rendezett p -értékekre. Az a legnagyobb p -érték, amely még a meghatározott egyenes alatt helyezkedik el, lesz a szignifikáns p -érték küszöb [148].

Ez a megközelítés dinamikusabb, mert figyelembe veszi a tesztek számát és azok eredményeit is. A BH korrekció lehetővé teszi, hogy több ténylegesen szignifikáns eredményt tartsunk meg, miközben kontrollálja a hamis pozitív arányt, így nagyobb statisztikai teljesítményt biztosít.

7.8.3 Érvek a Benjamini-Hochberg módszer mellett

A Benjamini-Hochberg módszer előnye, hogy dinamikusabb és rugalmasabb, mint a Bonferroni-korrekció. Míg a Bonferroni-módszer túl szigorúvá válhat nagy áteresztésű vizsgálatok esetén, addig a BH korrekció lehetőséget ad arra, hogy több valós pozitív eredményt tartsunk meg, ezáltal növelve a vizsgálat hatékonyságát és értékét. A BH módszer alkalmazása különösen akkor ajánlott, amikor nagyszámú tesztet végzünk, mivel jobban képes kontrollálni a hamis felfedezési arányt, és realisabb képet ad a szignifikáns eredményekről.

7.9 Háromdimenziós modellekhez használható szerkezetek

A Kir 2.1 az AlphaFoldot leszámítva a legnagyobb részt tartalmazó szerkezete a 2xky, hiányzó része a membárnon belüli régió. A másik két elérhető szerkezet (1U4F és 2GIX) csak a citoplazmában szereplő régió rendezett részét tartalmazza.

Az emberben is fellelhető NMDA receptorokról szerencsére elég sok térszerkezeti modellt találunk, olyan értelemben is, hogy, közel lefedik a szekvencia legnagyobb és lényegi részét, egyedül a terminálisokat leszámítva. Ezeket elektronmikroszkópiával készítették eltérő felbontásban 4,5 Å-től egészen 7,8 Å-ig.

Az AMPA receptor (AMPA) egy ionotrop glutamát receptor, amely alapvető szerepet játszik a gyors szinaptikus átvitelben az agyban. Az AMPAR egy tetramer fehérjekomplex, amely négy alegységből áll: GluA1, GluA2, GluA3 és GluA4. Ezek az alegységek különböző kombinációkban állhatnak össze, ami különböző funkcionális és farmakológiai tulajdonságokat eredményez [149], [150].

Az extracelluláris domén tartalmazza a ligandkötő helyeket, ahol a glutamát molekula kötődik. Ez a kötés indukálja az ioncsatorna nyitását, lehetővé téve a nátrium- és kalciumionok áramlását a sejtbe [150], [151]. Az N-term domén fontos szerepet játszik az alegységek közötti interakciókban és a receptor szerkezeti stabilitásában [150]. Az AMPAR négy transzmembrán szegmensből áll (M1-M4), amelyek az ioncsatorna pórusát alkotják. Az M2 szegmens intracelluláris hurok formájában található, és fontos szerepet játszik az ion permeabilitás szabályozásában [150]. Az intracelluláris C-terminális domén különbö-

ző fehérje-interakciós helyeket tartalmaz, amelyek szabályozzák a receptor traffickingját, foszforilációját és a szinaptikus plaszticitást [151].

Az AMPAR alegységek variabilitása hozzájárul a receptorok diverzitásához, amely lehetővé teszi az agy különböző területein és különböző funkcionális kontextusokban való működést. Például, a GluA2 alegység jelenléte befolyásolja a kalcium-permeabilitást, mivel a GluA2 tartalmazó AMPAR-ok általában nem permeábilisak a kalciumionokra, míg a GluA2 hiányzó AMPAR-ok kalciumot is átengedhetnek, ami jelentős hatással lehet a sejtélettani folyamatokra [151].

Az AMPAR-ok struktúrája és működése továbbra is intenzív kutatások tárgya, mivel kulcsszerepük van az idegrendszer normál működésében és különböző neurológiai betegségek patogenezisében [150], [151]. Az AMPAR szerkezetét számos PDB adatbázisban található szerkezet írja le. Néhány példa:

- GluA2 ligandkötő domén: PDB ID **3KG2** [152]
- Teljes AMPAR tetramer: PDB ID **5WEO** [153]

A PSD-95 emberi fehérjéből származó térszerkezetek nagyon hiányosak leginkább csak a PDZ doméneket tartalmazzák. Patkányból viszont már szerepelnek olyan térszerkezetek, amelyek a többi domént is tartalmazzák. A teljes fehérjét tartalmazó 2xkx kódú modell számos hibával terhelt, és most már maga a pdb adatbázis is szinte a teljes szerkezetet mismatchként jelöli meg.

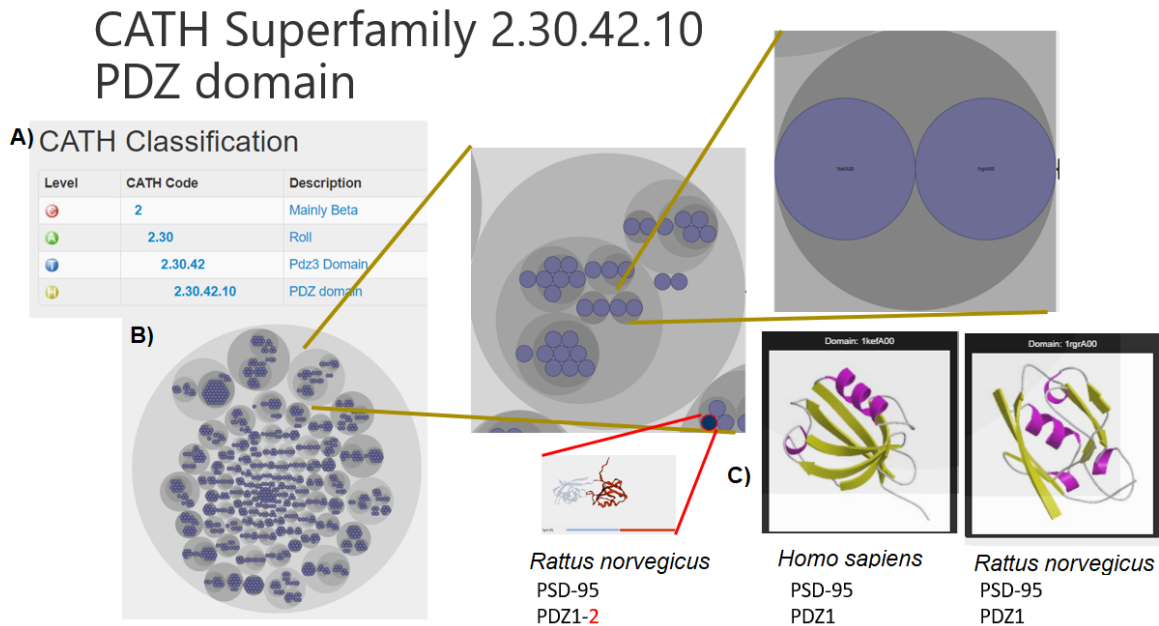
Részleges térszerkezeteket viszont találunk többet is a doménekről, a 7.6 bemutatja a CATH adatbázisban található PDZ domén családot és azon belül a PSD-95 PDZ1 doménjének elhelyezkedését. Látható, hogy a *Homo sapiens*ben és a *Rattus norvegicus*ban található PDZ1 domének a legkisebb klaszter. A mellettük levő klaszterben már Dlg3 PDZ1 doméneket találtam. A tandem szerkezeteket a mellettük levő nagyobb klaszterben találtam meg a PDZ2 szerkezet miatt.

Ehhez hasonlóan az emberben található PSD-95 fehérjére indított keresés a DisProt adatbázisban nem hozott találatot. Ellenben a *Rattus norvegicus*ban található PSD-95 fehérjére egyetlen egy találatot kaptunk ami szerint 1,52%-ban tartalmaz rendezetlen régiót. Az annotáció szerint ez a rendezetlen régió a $\beta 8$ - $\beta 9$ hurokrégióban található (569-579 aminosavak) az aszimmetrikus egységben, méghozzá a guanylate kinase domén közel elején [154]. Ez a szakasz nagyon rövid, önállóan kiemelésre nem szorul, mindazonáltal a szerkezet modell felépítésénél ezt a rövid szakaszt is kezelhetem más felbontásban, a szálak minimális elmozgását megengedvén. A hozzá párosított PDB szerkezet: 1JXM. A cikk további információkkal lát el minket arról hogy az SH3 és GK domének között intermolekuláris interakciók lépnek föl β -redő segítségével, mely tovább finomíthatja a fehérjemodellünkről alkotott képet. A referált szerkezet ezt a két domént tartalmazza csupán ligandkötött állapotban.

A PDB adatbázisban az 5JXB kódú szerkezet tartalmazza az egér PSD-95 PDZ3 doménjét a SynGAP PBM doménjével komplexben. Komplexképződés szempontjából ezek a szerkezeti elemek kifejezetten fontosak lehetnek, viszont a SynGAP hiányzó részeire is szükségünk van ahhoz jól láthassuk a teljes fehérje térégenyét. Mindazonáltal sajnos az emberi fehérjéhez tartozó térszerkezetet nem találtam.

A GKAP foszforilációtól függő módon képes a PSD-95 GK doménjével kölcsönhatásba lépni a megfelelő komplexet az 5YPO kódú szerkezet tartalmazza.

A Shank3 fehérjének is kevés szerkezete található meg az adatbázisban. A 6CPK csak az SH3 domént tartalmazza. A 7C7I szerkezet a fehérje N-terminális 1-99 aminosavját fedi le és



7.6. ábra. A PSD-95 PDZ1 doménjének elhelyezkedése a PDZ szuper családon belül CATH felosztásban A) A táblázat leírja az osztályok pontos megnevezését. B) A kiemelt ábrák az osztályokon belüli alosztályokat szemléltetik, melyeken belül a PDZ helyezkedik el. C) A szerkezeti ábrák szemléltetik a hasonlóságát a szerkezeteknek, melyek alapján kerültek egymáshoz közeli osztályokba. Az ábra a szerző saját munkája.

ez az ún. SPN domén, amelyik nem mindig jekenek meg önállóan annotált doménként. Ras fehérjével alkot komplexet. A vándorpatkányból származó fehérjéhez (Q9JLU4) elérhető az N-terminális szakaszt is tartalmazó modell, (5G4X) és számos PDZ domént (5OVA), valamint pár SAM domént (2F3N) tartalmazó szerkezet is. Mindazonáltal a 800-tól 1500-ig tartó szakasz szerkezetéről nincs jelenleg atomi szintű modell, ez a szakasz feltehetőleg rendezetlen, esetleg fibrilláris jellegű szerkezeti elemeket tartalmaz. Valószínűleg kezelhető lesz flexibilis szakasként kis felbontással, amennyiben sikerül összekötni az adott doméneket. A rendezetlen régióra sokaságokat generálhatunk a DIPEND módszer segítségével[155]. A másik lehetőség, a meglévő AlphaFold [156] által predikált szerkezet felhasználása a háromdimenziós modellezés tekintetében. Ebben az esetben figyelembe kell venni, hogy számos régiója, főleg a rendezetlen régiók nagyon alacsony megbízhatósággal rendelkeznek. Csoportunkban az elmúlt időszakban történtek erőfeszítések egy kezdeti teljes Shank szerkezeti modell felépítésére is, melyhez a rendezetlen részeket saját módszerrel állítottuk elő.

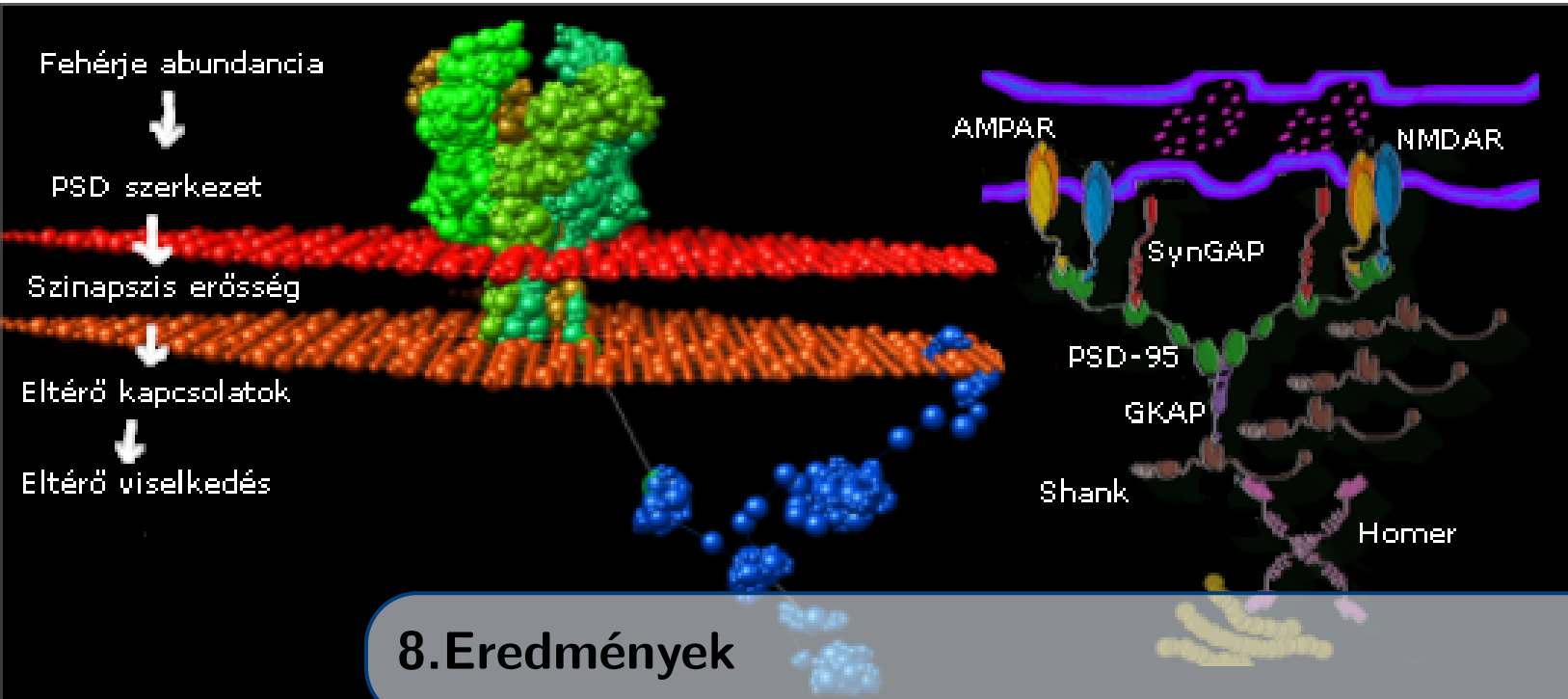
A Shank1 esetében több kísérleti szerkezet áll rendelkezésre, mindazonáltal a szerkezetekből 5 darab a Shank1 PDZ domént fed le, míg a 6CPI az SH3 domént tartalmazza. AlphaFold predikció a Shank1-ről is elérhető.

A Homer fehérje kapcsán sem elérhetőek az emberi fehérjére vonatkozó szerkezeti modellek, csak AlphaFold predikció. Egérből van pár szerkezet a fehérje N-terminálisáról nagyobbak és egy kisebb a C-terminálisról. Az N-terminális az EVH1 domént tartalmazza, míg a C-terminális szakasza tetramerizációhoz szükséges szakaszt foglalja magában. Így ezen

szerkezeteket is meg lehet próbálni összekötni és a kötéseket kisebb felbontásban ábrázolni, megengedve annak nagyobb elmozdulását mert a komplexek kialakításához ezek a régiók, amik mindenképpen nélkülözhetetlenek. Bár természetesen így is pontatlanabb eredményt kapnánk egy valós térszerkezethez képest. Ezért csak elsődleges modellezésre lesz alkalmas.

7.10 A szuperfelbontású mikroszkópia a szimulációk és modellek validálására

Szuperfelbontású mikroszkópia kifejezetten fontos új technológiai előrelépés a szubcelluláris világ megfigyelésében, hiszen a szubcelluláris világ nagyon kicsi, és az emberek alig látják, annak ellenére, hogy ez a világ óriási hatással van az emberi viselkedésre. A szuperfelbontású mikroszkópia az új módszer, amely "lehetővé teszi a szubcelluláris szerkezet megjelenítését a példátlan részletességgel", azaz az elvégzett szimulációk, majd a belőlük készített modellek részben ellenőrizhetővé váltak, részben pedig segítséget nyújtanak célzott mikroszkópiás vizsgálatokhoz. A szuperfelbontású mikroszkópia nem a mikroszkópia egyik típusa, inkább sokféle megközelítés esernyő fogalma, így összegyűjti a számos mérési eszközt és alapvetet. A Schermelleh et al. cikke [157] igen fontos eredményt mutat be, mert az intracelluláris protein képződmények ellenőrzése az emberi szem által kritikus és nehéz kihívás volt, de most pár évtized leforgása alatt a fehérjéket vizuálisan is meg tudtuk jeleníteni. A cikk nagyszerű áttekintést nyújt hét különféle szuperrezolúciós mikroszkópiás módszerről, és döntési fát nyújt ahhoz, hogy kiválaszthassuk kutatásaink számára a legmegfelelőbbet.



8. Eredmények

8.1 Az ágens-alapú szimuláció bemeneti modelljének kidolgozása

Az irodalom alapján kiválasztottam azt a hét kulcsfontosságú fehérjét, melyek egy összefüggő komplexet alkotnak, és ismereteink szerint befolyásolják a szinapszisok megerősödését, gyengülését - ezáltal közvetlenül dolgoznak az agy plaszticitásának megtartásán. Ez a hét fehérje a bevezetőben már bemutatott NMDAR, AMPAR, SynGAP, PSD-95, GKAP, Shank3, Homer1.

Következő lépésként olyan adatkészletet kerestem, amelyben ezen fehérjék lehetőleg minél több különböző gyakorsági kombinációja vizsgálható. Megfelelő mennyiségű közvetlen adat ezen fehérjék gyakoriságáról azonban nem állt rendelkezésemre, ezért a megtalált mRNS expressziós szinteket kellett átkonvertálnom darabszámokra.

Az irodalomban RNA-Seq kísérletekből származó RPKM (reads per kilobase million) értékeket tartalmazó adatokat találtam a [30] publikációban.

Ezekből az értékekből úgy állítottam elő darabszámra vonatkozó adatokat, hogy a PSD-95 ismert darabszámához viszonyítottam őket az alábbiak szerint:

1. Kiszámítottam a PSD-95 medián RPKM értékét
2. Ezt az értéket (66,82 rpkm) megfeleltettem az irodalomban találató PSD-95 darabszámnak (300)[158]
3. A kapott arányt (300/66,82) használva az egyes fehérjék egyedi RPKM értékeiből - lineáris összefüggést feltételezve Madeline Farley PhD értekezése alapján [159] - gyakorisági értékeket becsültem.

Kezdetben Shank3-al dolgoztam, majd a mutációanalízisnél Shank1-et vizsgáltam. A Cytocasttel végzett szimulációk tekintetében a fehérje doménszerkezetében nem jelent változtatást a domének relatív pozíciója - ezért könnyedén kicserélhető a két fehérje. Az egyes fehérjék becsült gyakorisága nagy eltérést mutat az egyes adatkészletekben. Az átlagos mennyiségeket a 8.1. táblázat foglalja össze:

A modell alkalmas elsődleges becslés készítésére. A fehérje mennyiségének mRNS expresszióval történő közelítése során figyelembe kell venni, hogy a régiók közötti eltéréseket

protein	min	max	átlag	szórás
NMDAR	0	95	16.84	17.87
AMPAR	1	688	126.81	78.39
PSD-95	36	1067	328.09	159.21
SynGAP	11	359	1102.21	60.13
GKAP	2	367	82.62	65.50
Shank1	1	388	69.20	59.40
Homer1	1	124	21.98	17.20

8.1. táblázat. A bemenetként használt fehérjegyakorisági adatok átlaga és szórása

már az mRNS szintézis szintjén megfigyelhetjük, azonban még a fehérjék translációja során más különbségek, lényeges csoportosulások is kialakulhatnak.

Valós fehérjebőségi adatok ismeretében az eljárás változtatás nélkül ugyanúgy megismételhető és pontosabb képet adhat. Ismét ki szeretném emelni, hogy ez a számítási mód csak egy közelítésnek tekinthető, számos helyen és formában modulálható a fehérje tényleges mennyisége expresszió után is.

Tekintve, hogy az *in silico* számítások a bemeneti adatoktól függenek, ezért nem elhagyható ezek behatóbb vizsgálata.

Észrevételeim a kiugró bemeneti adatokat érintik, és további megfigyelésekre és következtetésekre vezettek kutatásomat tekintve. Ezek a következtetések érintik az adatbázis összeállításáról szóló meglátásaimat is.

8.2 A posztszinaptikus denzitás mennyiségi és minőségi modellezése

8.2.1 Fehérjekomplexek integratív modellezése

Integratív modellezés során bármely forrásból származó strukturális információt felhasználva előállítjuk az összes modell halmazát, amely összhangban áll a felhasznált információkkal [160]. Az agy viselkedésének integratív modellezése számos szinten megtörténhet, génektől kezdve szinapszishálózatokon át funkcionális és strukturális megfigyelésekig [161]. Én a posztszinaptikus denzitás fehérjekomplexeinek szintjén modelleztem. Az információk összehangolásának lehető legjobb kivitelezése érdekében kidolgoztam egy komplex eljárási folyamatot, mellyel az irodalmazástól és a kísérletes adatoktól eljuthatunk a fehérjekomplexek háromdimenziós térkitöltő modelljeinek felépítéséig. Ezen létrejött modellek segítségével a posztszinaptikus denzitás tulajdonságait tovább vizsgálhatjuk, esetlegesen a kísérletes adatokkal összevethetjük, hiszen az eljárás során számos ponton ellenőrzési lehetőségeket helyeztem el.

Az eljárás két, párhuzamosan futó részből áll, amelyeket később lehet integrálni egy közös modellben - eképpen a Kar két kutatócsoportjának munkáját is összeköti a folyamat. Az egyik irányvonal fehérjekomplexek kialakulásának nagyskálás szimulációja Gillespie algoritmussal, amely során expressziós adatokból és különböző kötési adatokból komplexeket generálunk, úgy hogy az adott fehérjék ágensként viselkednek és egy adott rendszerben diffúzió segítségével találkoznak. A találkozás során adott valószínűséggel összeállnak egy komplexszé, ezután a kisebb komplexek még nagyobb komplexekké, akár egészen nagy, ún. szuperkomplexek is kialakulhatnak. Ebben az irányvonalban a komplexek pontok és élek halmazaként azaz gráfként vannak reprezentálva, és az eredményünk egyfajta feltételes

valószínűségi becslés, mellyel megkaphatjuk, hogy adott komplex milyen valószínűséggel jöhet létre, feltéve, hogy adott mennyiségű fehérje áll rendelkezésünkre. Ezzel ki tudjuk emelni a leggyakoribb interakciókat és kapcsolatokat, amelyek segítségével eldönthetjük, mely komplexeket célszerű elsődlegesen vizsgálnunk.

Az eljárás másik irányvonala a térszerkezeteken és a szekvenciák annotációin alapszik és figyelembe veszi az egyes fehérjék mint alkotóelemek háromdimenziós jellegét. Ebben az esetben a fehérjékből gyöngysormodellt hozunk létre, ezáltal egy valódi kiterjedéssel rendelkező objektumként tekintünk rá. A fehérjemodelleket ezután Monte Carlo szimulációk segítségével komplexekké tudjuk összekötni, és a folyamat során megengedjük a flexibilis részek elmozdulását is. Az így kapott modellrendszerek segítségével az előző eljárásban kapott komplexek létrejöttének relevanciája is megvizsgálható, különös tekintettel a háromdimenziós elrendeződés adta esetleges korlátokra.

Tehát a két irányvonal összefűzésével egy sok aspektust figyelembe vevő, összetett modellt kaphatunk, melyet a végén kiértékelhetünk. Az eljárásnak nem elhanyagolható része az, hogy minden egyes lépés után van lehetőség validációra, hogy minél több hibát kiszűrhessek, és lehetőleg már a validáláson átesett információkat felhasználva léphessenek tovább. A koncepciót az 8.1 ábra szemlélteti, melyben minden lépés esetében feltüntettem, hogy körülbelül milyen típusú kérdésekre adhatunk választ.

Így az expressziós adatok tekintetében célszerű megnézni magát az irodalmat és külön vizsgálni, hogy már az adatokban vannak-e kiugró értékek és hogy azokat miképpen kezelték a kísérletes eljárás során. Érdeemes lehet azt is megfontolni, amire később kitérek, hogy az expressziós adatok nem feltétlenül tükrözik a fehérjék tényleges mennyiségét. Fontos annak is az eldöntése, hogy az egyes szimulációk eredményei közötti eltérések szignifikánsak-e, vagy csupán a szimuláció sztochasztikus jellegéből adódnak.

A térszerkezeti információkat és adatbázisokban elérhető modelleket is elemznünk és validálnunk szükséges felhasználásuk előtt. Ilyen tekintetben például fontos a funkcionálisan rendezetlen, fibrilláris és globuláris régiók elkülönítése, illetve a PDB adatbázisban elérhető kísérletesen meghatározott szerkezetek geometriai jellemzői, mint kötéshosszak, Ramachandran plotok, esetleges atomi ütközések stb. A legtöbb fehérje esetében igaz, hogy nagy mérete miatt a teljes molekula térszerkezete nem, csak egyes szakaszainak kísérletesen meghatározott struktúrája érhető el, emiatt szükséges a teljes fehérje feltérképezése. Itt megjegyezhető, hogy ma már léteznek az AlphaFold eljárás segítségével készített teljes modellek is, ezeket természetesen legalább ugyanolyan kritikusan kell elemezni, mint a kísérletes szerkezeteket. A kísérletesen meghatározott szerkezetek megbízhatósága is széles skálán szór. Ezért a szerkezeti modelljeink összerakása során minden információt figyelembe kell vennünk ahhoz, hogy megfelelő fehérje gyöngysormodellt készíthessünk.

Ezen gyöngysormodelleket is validálni kell, amely során figyelembe vehetünk különböző aminosavak térkitöltésétől kezdve egészen a teljes gyöngysor stabilitásáig számos információt. Ha rendelkezésünkre áll elektronsűrűségről térbeli adat - például egyrészecskés elektron kriomikroszkópiával készült térkép (cryo-EM) [162] - akkor ráilleszthetjük a modellt [163]. A PSD modularitását a cryo-EM vizsgálatok is jól mutatják [164]. Miután megalkottuk a gyöngysormodellt a teljes komplexre, akkor szintén figyelembe kell vennünk az egyes potenciális interakciók során felmerülő sztérikus gátlásokat, ill. az ezekből adódó esetleges kompetíciót a kötőhelyek között. Ebből nyert információk visszacsatolhatók a gráfként kezelt szimulációkra - bizonyos komplexeket a Gillespie szimulációban kizárhatunk. Emellett távolságokat és energiákat minimalizálhatunk a komplex modell esetében is és azok alapján

újra definiálhatjuk magát a komplex modellt. Ezen túlmenően a komplexek elemzése a fehérje gyöngysormodelljére vonatkozóan is adhat visszacsatolást mert, a Monte Carlo algoritmus során jobban vizsgálhatóak a fehérje szabadsági fokai és esetleges következtetéseket levonhatunk a modellünk helyességéről, pl. hogy adott szakasz flexibilitását megfelelően modelleztük-e.

Természetesen ahhoz, hogy megbízhatónak tekinthető következtetéseket vonjunk le, a teljes folyamatot többször meg kellhet ismételnünk a korábbi tapasztalatainkat beépítve.

A fentebb vázolt eljárás egyes részei természetesen már korábban leírt és jelenleg is alkalmazott lépések, az egész folyamat és a kétféle modellezési megközelítés összekapcsolása azonban tudomásom szerint ebben a formában nincs még leírva a saját közleményeimet nem számítva.

Összességében az újdonságérték a két módszer vegyítése és a két kutatócsoport eredményeinek közös felhasználása.

Az ágens-alapú szimulációk esetében az egyik fő kérdés az, hogy hogyan tudjuk elemezni az input és output adatait a szimulációknak, ahhoz, hogy az felhasználható legyen a későbbiekben arra, hogy a kigondolt munkamenetben tovább lehessen haladni a háromdimenziós térszerkezeti elemzések felé. A szimulációk előzetes képet adnak a lehetséges fehérjekomplexekről kétdimenziós gráfszerű nézetben, mely nem feltétlenül jelenti azt, hogy az adott komplexek a fizikai háromdimenziós térben is létrejöhetnek, de a kapott komplexeloszlás jó kiindulópontot jelent a további elemzésekhez.

Az ágens-alapú szimulációkkal kapcsolatos kérdések:

1. mennyire robusztus a rendszer, a metaparaméterek mennyire befolyásolják az eredményeinket?
2. adott szimulációt hányszor szükséges megismételni, hogy a sztochaszticitás hatása kiküszöbölhető legyen?

8.3 A szimulációk reprodukálhatósága

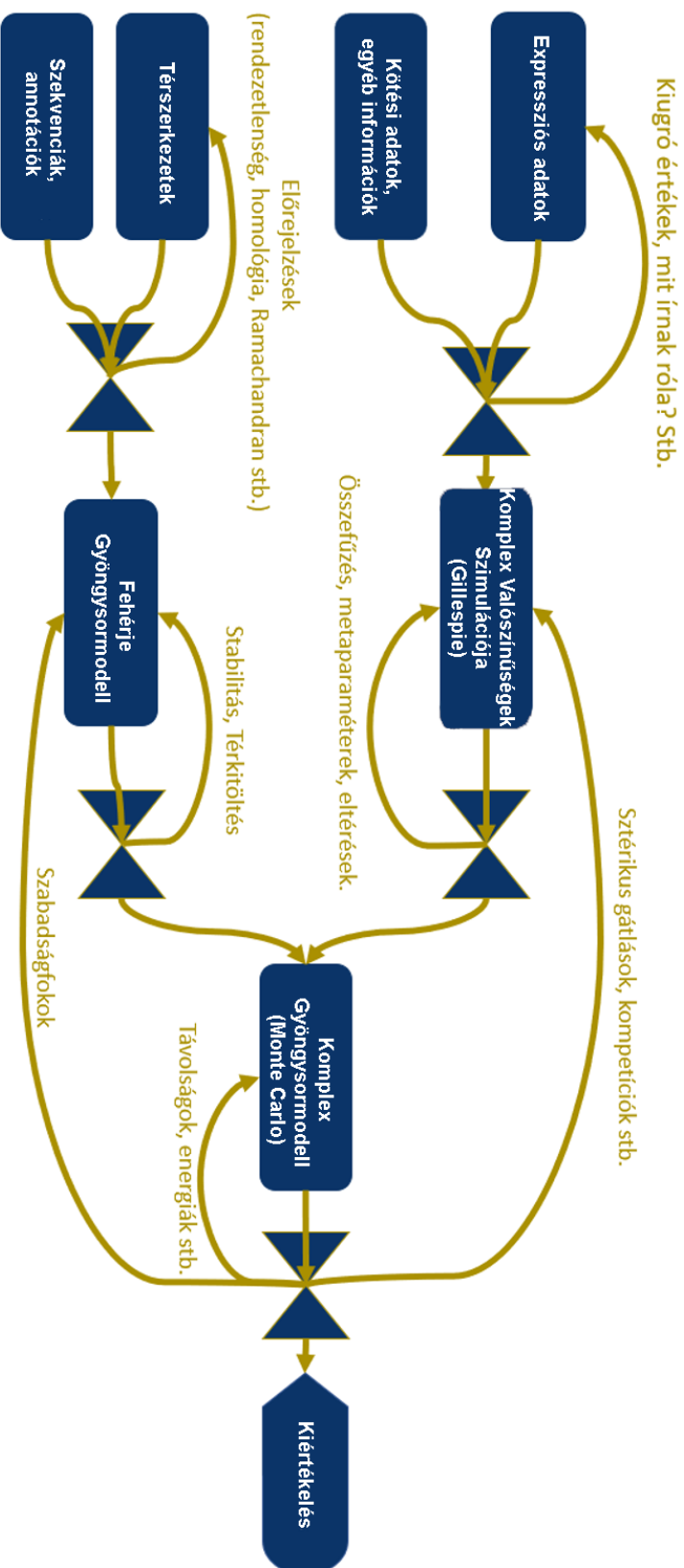
A szimulációkkal kapcsolatos egyik fő kérdés az, hogy mennyire robusztusak és megbízhatóak. Ahhoz, hogy meggyőződhessünk arról, hogy a szimulációs eredmények elfogadhatóak, számos alkalommal le kell futtatni őket ugyanazokkal a bemenetekkel. Ezen eredményeink átlagolása egy tapasztalati várható értéket ad, amely azonban még nem feltétlenül egyezik meg az ideális, determinisztikus várható értékkel, és hibája függ a kimenetként kapott komplex abundanciák szórásától és az ismételt szimulációk számától is.

A kérdés gyakorlati elemzéséhez ugyanazt a szimulációt n -szer futtatom, majd egy átlagos komplex darabszámot számítok ki. A konfidencia intervallum az az intervallum, amely 0,95 valószínűséggel tartalmazza ennek az átlagnak a környezetében az ideális várható értéket. A szimulációk ideális száma levezethető a konfidenciaintervallum sugarának szűkítésére.

$$\frac{\Phi^{-1}\left(\frac{1+p}{2}\right)\sigma}{r} < n \quad (8.1)$$

Ahol p a valószínűség, általában 0,95, σ a szórás, és r a kívánt konfidenciaintervallum sugara - $\Phi(x)$ a standard normál eloszlás.

Esetünkben a nehézség abban rejlik, hogy ezt az intervallumot minden komplexre ki kell számítani, így ugyanazon bemenetre a szimulációk optimális számát kell megtalálni. Ezért



8.1. ábra. Kutatási folyamatábra Az ábra a szerző saját munkája

számos ismétléssel (5,10,40,50,100) futtattam le a szimulációkat. Azt tapasztaltam, hogy közel 40 szimuláció már elegendőnek bizonyult ahhoz, hogy a legtöbb komplex esetében a szórás minimális legyen.

Összességében 524 különböző agyi régióhoz tartozó adatot használtam fel a szimulációk során [30]. Valójában az 524 adatsor 42 alanyból származik, akiknek több – nem mindig ugyanolyan módon definiált – agyterületéről is származhatott minta, ez teljesen alanyfüggő volt. Az alábbiakban az egyszerűség kedvéért ezekre az adatokra „agy régió”-ként vagy egyszerűen „régió”-ként fogok hivatkozni.

Az agyi régiókat 27 régiótípusba lehet sorolni, amelyekből az adatok származnak. Mind az 524 agyterületre 40 szimulációt futtattam.

8.3.1 A szimulációs idő hatása a komplexek méretére

A szimulációk eredményeit általában a viszonylag kis, 2-4 fehérjéből álló komplexek dominálják. A molekulák közötti potenciális kölcsönhatások azonban sokkal nagyobb asszociációk kialakulását is lehetővé teszik. A kisebb komplexek nagyobb komplexekké állhatnak össze, ezért külön megvizsgáltam, hogyan fejlődnek a nagy komplexek a szimulációim során, és kialakulásuk hogyan függ a szimulációk hosszától.

A Cytocast programmal számos hosszú, több napos szimulációt futtattam le, különböző szimulációs időket beállítva. Ez a metaparaméter azért fontos, mert, ha még a stabil állapot előtt leállítjuk a szimulációt, akkor tévesen hihetjük azt, hogy az adott komplexből nem keletkezett számottevő mennyiség. Ezért adott szimuláció-sorozatot különböző szimulációs időkkel is lefuttattam, és azt kaptam, hogy az eredetileg beállított szimulációs idő háromszorosa mellett sem változtak meg számottevően az eredmények, így a beállított „alap” szimulációs idő (nem tényleges futási idő, a tényleges futási idő három vagy több napot is igénybe vehet, addig a gépnek zavartalanul üzemelnie kell) is elegendőnek bizonyult ahhoz, hogy az eredményeket elfogadjuk.

Megjegyzendő, hogy a Cytocastban megvalósított Gillespie-algoritmus természetéből adódóan a szimulációs idő nem lineárisan skálázódik a lépések számával. Az algoritmus minden lépésnél kiszámítja, hogy mennyi ideig tart a következő reakció befejezése [165] - lásd a módszerekben. Emiatt a szimulációs idő és a lépések száma közötti összefüggés nem pontosan lineáris, hanem zajosabb jellemzőket mutat. Ennek eredményeként a duplikált szimulációs idő nem feltétlenül jelent pontosan megkettőzött lépésszámot.

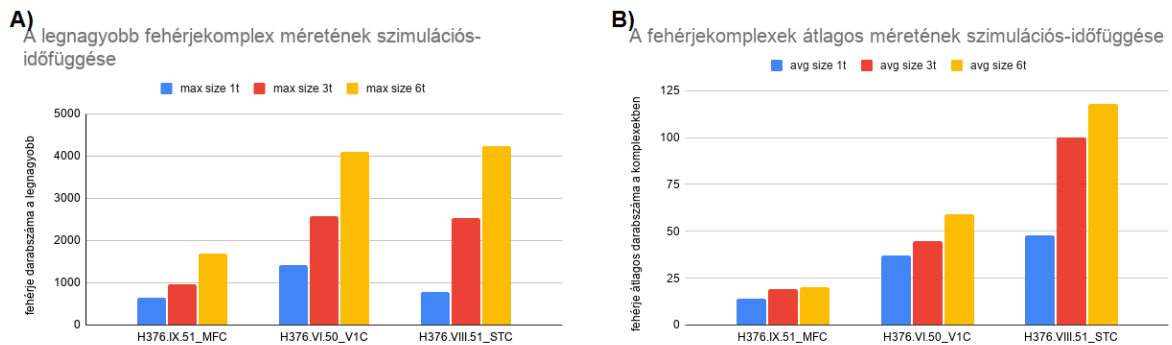
A szimulációs idő befolyását minden egyéb paraméter módosítása során is érdemes megvizsgálni, mert lehet, hogy egy újabb kötési lehetőség már kimozdítja a rendszert és a stabil egyensúlyt későbbre tolja ki. Ezek megfontolás alapján teszteket végeztem és megállapítottam, hogy közel stabil és hasonló eredményeket kapok a legtöbb beállítás során, így viszonylag hamar beáll a rendszer egy egyensúlyi állapotba elsősorban az elsődleges építőelemek arányától függően.

Elsődleges komplexeknek a tíz kötésnél kevesebb kötést tartalmazó komplexeket tekintettem, melyek a fő építőelemként bolyonganak a rendszerben, hogy később egyre nagyobb és nagyobb komplexekké kapcsolódjanak, és kialakítsanak egy vagy két szuperkomplexet, amely a posztzinaptikus denzitás területének nagyobb részét foglalhatja el. A legfőbb eltérés a kevesebb kötést tartalmazó elsődleges komplexek és a náluk jóval többet tartalmazó komplexek között a vártnak megfelelően az volt, hogy darabszámban lényegesen több elsődleges komplex jelenik meg minden szimulációban, mint az azokból létrejövő összetettebb nagyobb komplexek. Ez a jelenség egyszerűen abból adódik, hogy minél több fehérjéből áll

egy komplex, azaz minél több kapcsolatot kell létesíteni, kapcsolatok létesítésének együttes valószínűsége egyre alacsonyabbá válik. Ez természetesen *a priori* nem új információ, és alapvetően megbecsülhető a kapcsolatok számából és az affinitási értékekből. A megjelenési valószínűség a kötőszámmal fordítottan, az affinitás értékekkel egyenesen arányos.

Mivel az egyes fehérjék mérete nagy, ezért egy-egy komplex esetében egyetlen plusz fehérje bekötése is jelentősebb különbségekhez vezethet. Emiatt is érdemes a szimuláció során ezen szuperkomplexeket külön kezelni, és adott esetben a térszerkezeti modellezés során a kérdésre visszatérni, és megvizsgálni, hogy okoz-e jelentősebb térszerkezeti vagy funkcionális eltérést az adott fehérje jelenléte a kérdéses szuperkomplexben (pl. elviekben előfordulhat, hogy a helyzete miatt a „kifelé” elérhető kötőhelyek száma és jellege nem változik a teljes komplexeket tekintve).

Amennyiben várhatóan a két komplex nagyrészt ekvivalens lehet, akkor a későbbiekben érdemes lehet őket egységesen kezelni a kiértékeléskor, viszont alapesetben külön kell kezelnünk ezeket még akkor is, ha főbb jellegeiket tekintve közel állnak egymáshoz (a komplexeket reprezentáló gráfban pedig nyilván szomszédosak, ha egyetlen komponensben térnek csak el). Ez a jelenség is példa a kétféle modellezési megközelítés közötti kapcsolatra és visszacsatolásra.



8.2. ábra. **A szuperkomplexek szimulációs-időtől való függése** t alatt az egységnyi időfutást értjük, melyet a Gillespie algoritmus ad meg és a szoftver sajátja. Tehát egyszeres, háromszoros és hatszoros futásidőket nézünk. A) A legnagyobb fehérjekomplexek méretének szimulációs-idő függése. B) A fehérjekomplexek átlagos szimulációs-idő függése. Az ábra a szerző saját munkája

A szuperkomplexek mérete a szimulációs idővel ugyan változik (lásd 8.2. ábra), de a főbb építőelemek aránya nem, ez arra engedett következtetni, hogy ugyanolyan arányban épülnek be a szuperkomplexekbe az elsődleges építőelemek – ezáltal az elsődleges komplexek gyakoriságának aránya indikátor jellegként elfogadható adott régió esetében a szimulációs időtől függetlenül. Amennyiben két elsődleges komplex aránya jóval nagyobb a többi elsődleges komplexhez képest, akkor nagyobb valószínűséggel várható, hogy nagyobb mennyiségben van jelen a két komplexből összeálló nagyobb komplex is. Ez többek között a két elsődleges komplexet összekötő kötés affinitásán múlik, valamint azon, hogy a két elsődleges komplex térszerkezete továbbra is megengedi-e annak az adott interakciónak a létrejöttét. Ezért az ilyen számításokat is a háromdimenziós térszerkezetek meghatározása után érdemes újraértékelni, amikor már látható, hogy a térszerkezet hogyan befolyásolja

az adott kölcsönhatást. Amennyiben a kötést akadályozza valamelyik elsődleges komplex térszerkezete, akkor a két komplexből összeálló nagyobb komplexnek sem lenne szabad a szimulációban megjelennie ahhoz, hogy hitelesebb képet kapjunk. Az ilyen visszacsatolások lehetőségét szintén érdemes a használt szimulációs programokba beépíteni.

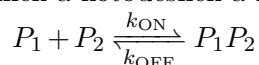
Fontos további aspektus, hogy a kialakult szuperkomplexek jellemzően egyediek, és nagyon alacsony, általában 1 körüli példányszámban vannak jelen. Továbbra is igaznak tűnik, hogy a leggyakrabban előforduló kisebb fehérjeasszociátumok száma és eloszlása nem változik jelentősen, bár a fehérjék állandóan, dinamikusan cserélődnek a komplexek között. A komplex méretének súlyozott átlaga a szimulációs időtől függetlenül a 2,7-2,8 tartományban marad, ami azt jelenti, hogy a bináris és háromkomponensű komplexek maradnak meg a legnagyobb mennyiségben.

A hosszabb szimulációs idő nem változtat érdemben a legnagyobb mennyiségben előforduló komplexek előfordulásán, amelyek valójában sokkal kisebbek, mint a nagy, nagyon alacsony példányszámú komplexek. A kisméretű, gyakoribb komplexek az interakciókon keresztül létrejövő nagyobbak építőköveinek tekinthetők. Így a szimulációs idő úgy hat a legnagyobb komplexekre, hogy minél hosszabb a szimuláció, annál több kis komplex áll össze szuperkomplexekké anélkül, hogy a rendszer kisebb blokkjainak dinamikai egyensúlya megváltozna. A jelenség a fázisszeparációval is párhuzamba állítható lehet, bár expliciten ezt a szimulációnk nem tartalmazza.

8.4 A mutációk hatásainak kiszámítása kötéseerősség predikcióval

Munkám során adott mutáció hatását is vizsgálni kívántam, viszont adott mutációt tartalmazó fehérjére (ill. fehérjedoménre) vonatkozó kísérleti kötési adat nagyon ritkán áll csak rendelkezésünkre, ráadásul a Gillespie algoritmus külön kezeli az asszociációs és disszociációs rátákat, amelyekre vonatkozóan vad típusú fehérjék esetében is alig van adatunk. Egy mutáció partnerkötésre vonatkozó közvetlen hatása nem könnyen becsülhető, viszont több olyan módszer is elérhető, amelyek adott fehérje stabilitásának megváltozását becslik, így ebből kiindulva próbáltam meg a kötési állandókat módosítani a szimulációs rendszerben.

Modellemben komplexképződés történik, ezért a komplexképződés leírható egy olyan reakcióval, ahol két fehérje vagy komplex kölcsönhatásba lép és alkot egy nagyobb komplexet. Ennek a kötődésnek a képződését és lebomlását az egyenlet együtthatói paraméterezik.



A szimuláció során fehérje degradáció nem valósul meg. Tehát a mutáció által bekövetkező stabilitásváltozást máshogyan kell megjeleníteni a szimulációban. A fehérje stabilitásvesztésének, (részleges) kitekeredésének lehetőségét a rendszerben a komplexképződés valószínűségének módosításával érem el.

A pontmutáció megváltoztathatja a fehérje stabilitását, ezért módosítja a fehérje hozzáférhetőségét. A biológiailag releváns pontmutációk stabilitásváltozását ($\Delta\Delta G$) a NeEMO [137] segítségével számítottam ki.

A kötés disszociáció aránya ($K = \frac{[\text{denaturated}]}{[\text{naturalated}]}$) a $\Delta\Delta G$ -ból származtatható. A változási arány azt mutatja meg, hogy a disszociációs ráta hogyan változik a mutáció által a megjósolt $\Delta\Delta G$ értékeknek megfelelően.

$$\Delta\Delta G = -RT \ln(K_{\text{WT}}) + RT \ln(K_{\text{MT}}) \quad (8.2)$$

$$\text{változási_arány} = \frac{K_{\text{MT}}}{K_{\text{WT}}} = e^{\frac{\Delta\Delta G}{RT}} \quad (8.3)$$

A modellemben a komplexképződés valószínűsége fordítottan arányos a disszociációs ráta (K) változásával. A disszociációs állandót módosítottam.

$$k_{\text{off_MT}} = k_{\text{off_WT}} e^{\frac{\Delta\Delta G}{RT}} \quad (8.4)$$

8.4.1 A Shank1 R743H mutáció kiválasztása és modellezése

A specifikus modellezési célunkra a Shank1-ban található R743H mutációra esett a választásom. Bár a COSMIC[166] adatbázisból választottam ki, amely rákos megbetegedésekkel kapcsolatos mutációkat tartalmaz, szándékom az volt, hogy egy jól leírt mutációt használjunk egy olyan globuláris tartományban, amely egy, a modellünkben szereplő és szerkezeteileg jól karakterizált interakcióra lehet hatással. Esetünkben ez a Shank1 PDZ doménje és a GKAP C-terminális régiója közötti kapcsolat. A mutáció komplexképződésre gyakorolt hatását a fentiek szerint a PDZ domén destabilizálódásából becsültem. A destabilizáció mértékét a NeEMO [137] eljárással becsültem.

Az R743H mutációt a PDZ domént érintő mutációk sorozatából választottam ki, mivel ez mutatta a legnagyobb destabilizáló hatást. Becslésem szerint a mutáció a Shank1:GKAP komplex disszociációs sebességének 5,5-szeres növekedését okozza.

Ez az 5,5-szeres értéket jó kompromisszumnak tekintjük a minimálisan megfigyelhető 2-szeres változás és az érték egy nagyságrenddel történő megváltoztatása között, ami egy erőteljesebb változás lenne. A mutáció által érintett arginin a 2. hélix C-terminális végén található, míg az R736Q mutáció, amelyet az ASD-vel [104] kapcsolatban írtak le, ugyanazon hélix N-terminálisát érinti. Mindkét arginin a közvetlen ligandumkötő helytől elfelé mutat, így várhatóan mindkét mutáció közvetett módon megzavarja a kötési kölcsönhatást.

8.5 Az ágens-alapú szimulációk értékeléséhez kidolgozott eljárások

Az adatok eloszlása nem egységes abból a szempontból, hogy vannak olyan agyi régiók, amelyekre vonatkozóan nem állnak rendelkezésre adatok az összes betegről. Emiatt minden kísérletet külön-külön kell kezelni, és arra törekedtem, hogy azonosítsam az eredmények közötti főbb különbségeket.

8.5.1 A keletkezett komplexek gráfszerkezetének vizsgálata

Homomultimer fehérjeasszociátumok azonosítására egy külön algoritmust hoztam létre. Erre azért volt szükség, hogy a nagyobb komplexek esetében is átfogó képet kapjunk arról, hogy az adott multimerizálódni képes fehérjék hogyan, s miképpen ragasztották össze az akár 100-as fehérjeszámot elérő komplexeket. Fontosnak tartom leszögezni, hogy attól függetlenül, hogy szimulációimban kijöhetnek szuperkomplexek, melyeket nagyon sok összeragadt Shank fehérje tesz lehetővé - közel sem biztos, hogy a természetes környezetében is megtörténik - gondoljunk csak a fázisszeparáció létrejöttére, amely szimulációimban is megjelenik.

Az azonos fehérjéből álló fehérjeasszociátumok megtalálásakor elsősorban meg kell határozni, mely fehérje képes multimerizálódni láncszerűen. Ez a kezdeti konstrukcióból könnyen

felismerhető. Az én tesztrendszeremben pl. a Shank3 fehérje kapott ilyen tulajdonságot az irodalomban leírt, SAM doménre jellemző asszociációja miatt [101].

Így egyszerűen kinyerhető, mely fehérjék játszhatnak adott komplexben szerepet a láncokban. Ezen fehérjéket az egyedi azonosítójuk (ID) alapján meg lehet keresni a fehérjekomplexekben és azon interakciókat azonosítani, amelyekben mindkét, egymással kölcsönhatásba lépő fehérje az adott proteinek felel meg.

Ezek után az eljárás végigveszi az azonosított fehérjepárokat, és megvizsgálja, hogy a pár valamelyik eleme szerepel-e már egy nagyobb asszociátumban („lánc”). Ha szerepel, akkor a pár másik, a láncban még nem szereplő elemét is hozzá kell adnia lánchoz. Amennyiben nem szerepel, akkor új láncot kell létrehozni.

Amennyiben a két pár között más pároknak is lennie kell a láncban, de még nem került a köztes párokra sor, egy új lánc keletkezik - ezt a láncot később kell összefűzni azzal a láncsal, amihez eredetileg kapcsolódott volna.

Az eljárás hasonlít a Kruskal algoritmusra [167] abban a tekintetben, hogy mindkettő addig választja ki az éleket, amíg lehetséges, és ezek az élek csatlakoznak a már kialakuló láncokhoz, ha egy láncot alkotnak. A legnagyobb különbség az, hogy ebben az eljárásban több különálló fehérjelánc is kialakulhat, nem alakul ki feszítőfa, és bizonyos esetekben akár kört is tartalmazhatnak.

Shank láncok meghatározása után megvizsgálható, hogy ezen láncokhoz kapcsolódó fehérjék milyen eloszlást mutatnak, melyet egy adott komplex példáján szemléltettek a 8.3 ábrán.

8.5.2 A komplexek azonosítása

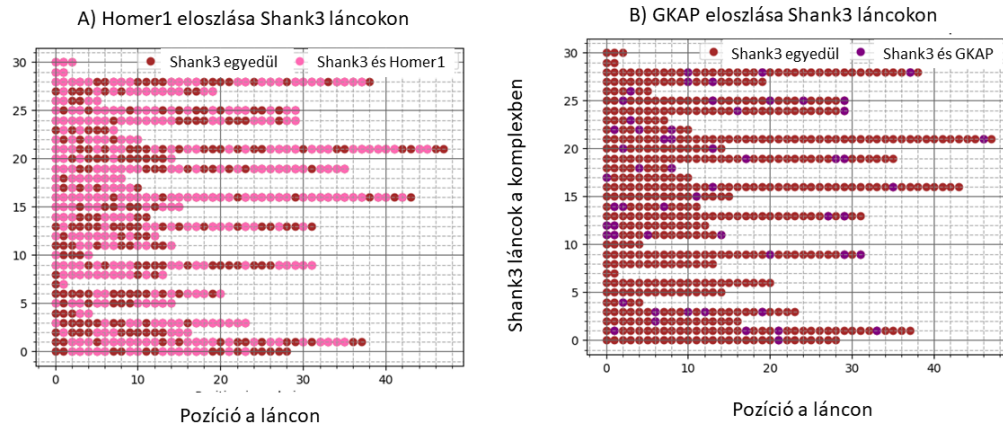
Szükséges az összes komplex mennyiségi összehasonlításának módszerét meghatározni. Ez kihívást jelent, mivel a komplexek általában több mint 10 alkotó fehérjét tartalmaznak, és nem csak fehérjetartalmukat szeretnénk összehasonlítani a komplexeknek, hanem a teljes szerkezetükre, azaz az egyes fehérjék kapcsolódási módjára is kíváncsiak vagyunk. Azon túl, hogy a polimerizált fehérjeláncokat figyelembe vesszük, fontos az is, hogy egyértelműen azonosítani tudjuk – a kapcsolódások figyelembevételével is –, hogy mely komplexek azonosak, illetve a hasonlóságukat definálni tudjuk.

A szerkezet tulajdonképpen leírható egy színezett gráffal, melynek minden csúcsa egy adott fehérje, élei pedig a kötések. A csúcsok színei az adott fehérjetípust reprezentálják, például zöld színű csúcsok a PSD-95 fehérjéket, míg a piros színű csúcsok a SynGAP fehérjéket jelölik. (Nem az számít, hogy konkrétan melyik fehérje köt adott pozícióba, hanem az, hogy az a fehérje AMPAR, vagy NMDAR.) A gráfot az is bonyolítja, hogy az egyes kötőhelyek szabadok és foglaltak is lehetnek, és adott esetben ezt is reprezentálni szeretnénk.

Bizonyos esetekben egy komplexben a fehérjék száma bőven meghaladhatja akár a 20 darabot is. Ezeket a komplexeket tekintem szuperkomplexeknek. A szimulációk többségében ilyen szuperkomplexek is keletkeznek, és általában csak kevés példányban található meg.

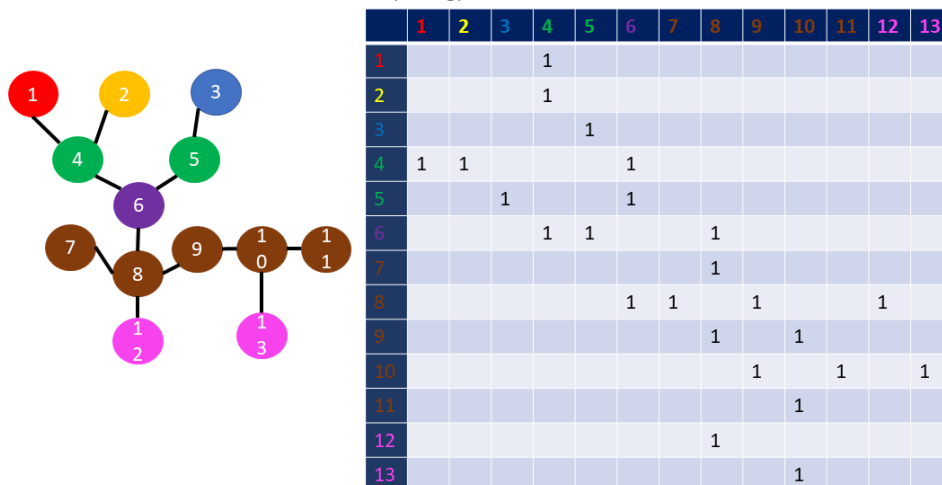
Az általam két komplex azonoságának megállapítására használt algoritmus diszkrét matematikai megfontolásokon alapul. Az eljárás az alábbi lépésekből áll:

1. Ellenőrzi, hogy a két komplex esetében azonos-e a fehérjekomponenseket leíró vektor, mely a komponensek ABC-sorrendben rendezett listája, a komplexben többször előforduló fehérjék többször szerepelnek ebben a reprezentációban.
2. Létrehozza a szomszédsági mátrixokat lásd 8.4 .



8.3. ábra. **Homer1 és GKAP eloszlások Shank3 fehérjeláncokon egy adott felhalmozódott superkomplexben** Az ábrán egy sor egy adott polimerizálódott Shank-lánc adott komplexen. A példában 30 különálló Shank-lánc szerepel a komplexben. A Shankok képesek kötődni Homerhez és GKAP-hoz. Ezen kötések eloszlása látható. Az ábra a szerző saját munkája

Az ID-k adott színben belül felcserélhetők, a komplex ugyanaz marad.



Lehetséges szomszédsági mátrixok száma: $1! \cdot 1! \cdot 1! \cdot 2! \cdot 1! \cdot 5! \cdot 2! = 480$

8.4. ábra. **Lehetséges szomszédsági mátrix adott komplex esetén.** A színek adott fehérjét jelentenek, például a barna a Shank. Valójában a fehérje számát, nem a fehérjéhez tartozó azonosítószám. Emiatt az adott komplexet leírni képes tényleges szomszédsági mátrixok száma az azonos színű (típusú) fehérjéktől függ. Az ábra a szerző saját munkája.

- Létrehozza a szomszédsági mátrixok minden permutációját csúcstípusok alapján. Nem triviális!
- Kiválaszt egy mátrixábrázolást az első komplexhez és ellenőrizze, hogy szerepel-e az egyes szomszédos mátrixok listájában a második komplex.

Az algoritmus legnagyobb hátránya, hogy faktoriális komplexitású, így a gyakorlatban futásideje miatt sajnos nem volt használható. Ehelyett több heurisztikát tartalmazó algoritmust generáltam, amely így is „túl szigorú” feltételeket szabott. Így már a minimális különbségek esetén is azt a választ adta, hogy különbözőek ezek a komplexek. Konceptciónk szerint minél nagyobb a komplex, azaz minél több fehérjét tartalmaz, annál kevésbé lehet releváns egy-egy komponens megléte vagy nem megléte. A heurisztikákat tartalmazó algoritmus rendszerspecifikus, így különböző lehetséges komplexösszeállítások során különböző heurisztikák vezethetők be, mindazonáltal ezen heurisztikák mintájára bármikor általánosíthatók azok.

Ilyen heurisztikának tekinthető például annak a vizsgálata, hogy egy szuperkomplexen több jellemző is definiálható, pl. a Shank3 láncok hossza, hány Shank3-GKAP kölcsönhatás van, milyen membránreceptorok kapcsolódnak a komplexhez és egymástól milyen távolságra stb..

A leírt megfontolások alapján komplex összehasonlítási algoritmusokat gondoltam végig, és ennek során arra jutottam, hogy az a legelőnyösebb, ha a problémát a Cytocaston belül kezelem. Ehhez minden összehasonlítandó régiót egy nagy, egyszerre futtatott szimulációban kell vizsgálni. Ekkor az egyes – a számítások szintjén különálló, de közös reprezentációval

rendelkező – régiókban az azonos komplexek egyedi azonosítója meg fog egyezni egymással, nagymértékben megkönnyítve az összehasonlítást. Ez azt is jelenti, hogy a mutáns és a vad típusú esetekben is közösen érdemes kezelni a futásokat. Ehhez úgy kellett kialakítanom a környezetet, hogy mind a 2^*524 régiót egyszerre megadhassam a szoftvernek.

8.5.3 A szimulációk hozzáadott információtartalmára utaló mérőszám meghatározása

Az általam elvégzett vizsgálatok egyik célja az volt, hogy megvizsgáljam, hogy a létrejövő fehérjekomplexek eloszlása – mint a szimulációk kimenete – hogyan függ az elérhető – a szimulációban bemenetként megjelenő – fehérjegyakoriságotól. A kérdés egyik általános vonatkozása, hogy a bemenetekből triviálisan következik-e a kimenet, vagy a kettő közötti összefüggés olyan bonyolult, hogy szimuláció nélkül nem lehet nagy biztonsággal becslést tenni adott bemenet esetén a kimenetre. Ennek eldöntésére megvizsgáltam, hogy a bemeneti és a kimeneti adatsorok ugyanazt a mintázatot adják-e az egyes régiókra (bemeneti adatkészletekre) vonatkoztatva.

Az összehasonlítást megnehezíti többek között, hogy a bemeneti és a kimeneti adatok más vektortérben helyezhetőek el – ezért közvetlen összehasonlításuk nem lehetséges. Ugyanakkor a bemenetekből és a kimenetekből is klaszterezhetőek. A bemeneteket a \mathbb{R}^7 hétdimenziós vektortérben lévő vektoroknak tekintettem. Az egyes dimenziókat, azaz a koordinátákat az adott fehérje gyakorisága határozza meg. Míg a kimeneteket kezdetben hatdimenziósként tekintettem \mathbb{R}^6 , ahol minden dimenziót az interakcióbőség alapján rendelttem hozzá, majd később egy nagyobb sokdimenziós módon vizsgáltam, ahol adott tengelyt adott komplex abundanciája jelentett \mathbb{R}^n , ahol az n az összességében keletkezett különböző komplexek száma.

A K-means clustering [141] eljárás az egymáshoz hasonló adatokból csoportokat hoz létre. A vektorterekben ez a tulajdonság azt jelenti, hogy a közelebbi pontok ugyanabba a klaszterbe kerülnek. Ezután minden klaszterhez létrehoztam egy-egy címkézett vektort. Az 524 hosszú vektor minden eleme egy vizsgált régiót reprezentál, és a koordináta 0, ha az adat nincs a klaszterben, és egy, ha az adat a klaszterben van (ún. one-hot reprezentáció). Ezt a one-hot reprezentációt arra használjuk, hogy két különböző klaszterezés esetén leírassuk, hogy az egyes klaszterek mennyire az egyes régiók ugyanolyan csoportosítását jelenítik meg, például bemeneti és kimeneti adatok alapján.

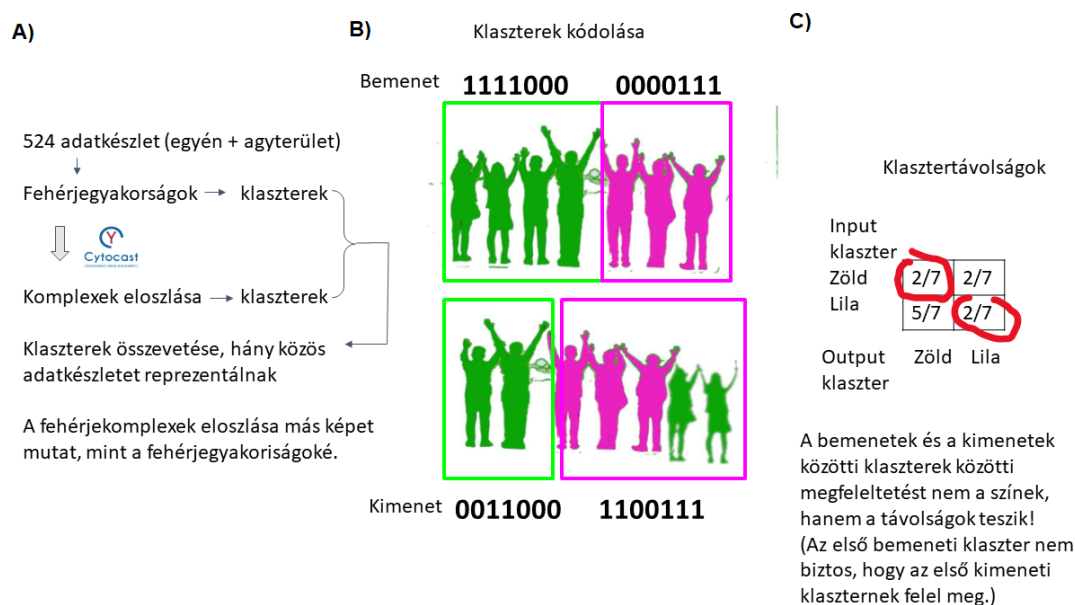
$$c \in \{0, 1\}^{\text{Adatpontok száma}} \quad (8.5)$$

$$c_i = \begin{cases} 1 & \text{ha } \text{adat}_i \in \text{Cluster} \\ 0 & \text{ha } \text{adat}_i \notin \text{Cluster} \end{cases} \quad (8.6)$$

Ekkor két klaszter távolsága formalizálható (lásd 8.5. ábra):

$$d(u, v) = \frac{\sum_{i=1}^N |u_i - v_i|}{N} \quad (8.7)$$

Ahol a u, v klasztervektorok, a N pedig az adatpontok száma (a klaszter vektortér mérete). Ez egy konzervatív megközelítés. Azokat a pontokat is hasonlóságnak tekintjük,



8.5. ábra. **Klaszterek kódolása és távolságuk** A) Klaszterezést el lehet végezni mind a bemeneten, mind a kimeneten, vagy mutációanalízis során a vad típusú kimeneten és a mutáns típusú kimeneten. B) Ezek után a klaszterek kódolhatóak attól függően, hogy az adott szimuláció benne van-e. C) Távolság definiálható a bemeneti és a kimeneti klaszterek között. Ez a távolság aképpen számítható, hogy hány bitben különböznek (például a két zöld klaszter 2 bitben különbözik egymástól.), majd az normálva van a bitek számával. A legalacsonyabb távolságú klaszterek a bemenetek és kimenetek között megfeleltethetők egymásnak. A megfeleltetett klasztertávolságok mutatják meg, hogy mennyire tértek el a klaszterek egymástól, mely klaszterekben lehetséges hozzáadott információ, amit ki lehet elemezni. Az ábra a szerző saját munkája.

amelyek nem szerepelnek a két klaszterben. Emiatt a kisebb értékek is nagyobb különbséget jelentenek, ha csak a két klaszter unióját vesszük figyelembe.

Ha a bemeneti és kimeneti klaszterek keresztávolsága elég nagy, akkor az összesített szimuláció olyan további információkat tartalmazhat, amelyek magukban a bemenetekben nem láthatóak. Az eljárás használható mind az interakcióbőség alapján definiált kisebb, hatdimenziós, mindpedig az egyes komplexek abundanciája alapján definiált sokdimenziós kimeneti komplexvektorteret használva is.

8.5.4 A legmagasabb információtartalmú komplex azonosítása

Az adatok értékelése során felmerült az igény arra, hogy meghatározzam azt a komplexet, amelyiknek a gyakorisága a legtöbb információt hordozza, azaz a leginkább hozzájárul a kapott fehérjekomplex-eloszlások közötti eltérésekhez. Ennek meghatározására egy főkomponens-elemzésen alapuló mérőszámot határoztam meg.

Az agyi régiókat egy többdimenziós térben lévő pontokként reprezentáltam, ahol minden koordináta egy adott fehérjekomplex gyakoriságát jelenti.

Egy adott agyi régió fehérjekomplex eloszlását lineáris kombinációként az alábbiak szerint határozzuk meg:

$$c \in \mathbb{R}^n, \quad c = \sum_{i=1}^n \alpha_i p_i \quad (8.8)$$

ahol az α_i az az i . komplex (p_i) gyakorisága (példányszáma), p_i egy alap (egység) vektor az i -edik tengely mentén, amely az i -edik fehérjekomplex másolatszámát jelenti.

Mindegyik komplex jellegzetes gyakorisági profillal rendelkezik a különböző régiókban: a legváltozatosabb abundanciájú komplexek mutatják a legszembetűnőbb különbségeket a régiók között. Így a legváltozóbb komplexek határozzák meg a rendszer fő összetevőit.

A főkomponens-elemzés egy hipotézismentes eljárás, melynek segítségével az eredeti vektorteret úgy transzformáljuk, hogy az egyes új tengelyek – melyek az eredeti vektortér tengelyeinek lineáris kombinációi – a vizsgált adatok varianciáját maximalizálják. Az első főkomponens felel a variancia legnagyobb hányadéért, az erre ortogonális második a következő legnagyobb hányadért és így tovább.

Az egyes komplexre vetített varianciát úgy számítjuk ki, hogy az egyes főkomponensvektorok által reprezentált varianciát megszorozzuk az illető komplex mint bázisvektor adott főkomponenshez való hozzájárulásával, majd ezeket minden főkomponensre összegezzük. A kapott relevancia megadja, hogy az adott komplex gyakorisága mennyire meghatározó az egyes régiók közötti különbségek meghatározásában.

$$r \in \mathbb{R}^n, \quad r = \sum_{i=1}^n \frac{\lambda_i}{\sum_{j=1}^n |v_{ij}|} v_i \quad (8.9)$$

Ahol r az egy n dimenziós vektor, ami tartalmazza az összes komplex relevanciáját, λ_i az i . sajátvektor relevanciája, v_{ij} a j . koordinátája az i . sajátvektornak (v_i).

8.5.5 Fehérjekomplex abundanciák összehasonlítása

Ahhoz, hogy kvantitatív állításokat tehessenek, szükség volt az egyes szimulációs kimenetek statisztikai összehasonlítására, a különbségek statisztikai szignifikanciájának megadására. Ez azonban a komplexek nagy száma miatt nem adható meg egyszerűen.

A páros t -próba egy klasszikus statisztikai próba, amelyet általában akkor választanak, ha csak egy változás - esetemben pl. a vizsgált Shank1 R743H mutáció - lehet jelen a rendszerben, és a kérdés az, hogy a változás hogyan befolyásolta az adott változó átlagértéket. A nullhipotézis: az átlagos abundanciák (vad típusú és mutáns) adott szignifikancia szint mellett azonosak. Így az alternatív hipotézis: egy komplex átlagos abundanciái jelentősen eltérnek. Egy adott komplexre a t -próbát a módszerekben ismertetett módon hajtottam végre.

Tekintettel arra, hogy a szimuláció során az egyes komplexek abundanciáját külön-külön átlagoljuk, a t -próba, hogy melyik hipotézist fogadhatjuk el, csak az adott komplex abundanciájára vonatkozik. Ahhoz, hogy egy adott régióról általánosságban meg tudjuk mondani, hogy változott-e, minden komplexen el kell végeznünk a t -próbát, és ezek súlyozott átlaga jelzi, hogy az adott régióra átlagosan melyik hipotézist tudjuk elfogadni. Az átlagolás

során az egyes komplexek ismert fontosságát használjuk fel, amelyet az egyes komplexek relevanciája alapján számítunk ki. A jellemző fontossága (feature importance) a t -próbával számított p -értékek átlagolásának súlya egy adott régióban minden komplexre.

$$\text{averaged p-value of a region} = \sum_{i=0}^{39} \text{complex_relevance}_i \cdot \text{p-value}_i \quad (8.10)$$

Adott régiótípusra átlagolt p -értékeket agyra vetítve lásd a [S1](#) ábrát.

8.6 A szimulációk hozzáadott információ tartalommal rendelkeztek

Kutatásom alapvető célja egy *in silico* eljárási folyamat kidolgozása volt integratív komplexmodellezésre. Ebből adódóan az egyik elsődleges kérdésem az volt, hogy a szimulációs lépés eredményei valóban tartalmazznak-e hozzáadott információkat, melyeket a talált fehérje gyakorisági adatokból a szimulációk lefuttatása nélkül nem vagy nehezen láthatnánk. Ez a fajta hozzáadott információ alapozza meg, hogy a folyamatba a Cytocast ténylegesen beilleszthető-e, vagy másképpen megfogalmazva, érdemes-e beilleszteni az ágens-alapú modellezést a munkafolyamatba.

Ennek a megvizsgálására matematikai modelleket dolgoztam ki (melyeket a már ismerttettem fentebb [8.5.3](#) részletesebben kifejtettem), hogy a be- és kimeneti adatsorok összehasonlíthatóvá válhassanak. Így különböző klasztertávolság számításokat is végrehajtottam saját Python kódok segítségével, amelyek a kidolgozott modellek alapján számolnak.

Az első megközelítés klaszterezésen alapult. Az alapelv, hogy azt vizsgálom, hogy a be- és a kimeneti adatsorok klaszterezése ugyanolyan módon csoportosítja-e a vizsgált régiókat. Amennyiben igen, úgy a kimenetben, azaz a kapott fehérjekomplexek eloszlásában nincs érdemileg új információ a bemenethez, vagyis a kiindulási fehérjegyakoriságokhoz képest. Az egyes klasztereket ún. one-hot kódolással reprezentáltam egy n elemű vektorban, ahol n a régiók száma, és az érték 1, ha a régió szerepel a klaszterben és 0, ha nem. Minden, a bemenetből származtatott klaszterhez megkerestem a leghasonlóbb, kimenetből számolt klasztert, ahol a hasonlósági metrika a közös elemek száma normálva a vektor hosszával. Az eljárást elvégeztem különböző számú klaszter esetére a K-means klaszterező eljárást használva. A legjobb megfelelést az $n = 3$ esetben kaptam, ahol a minimális klasztertávolságok 12%, 26% és 27% -nak adódtak, amik meglehetősen nagy értékek. Ez azt jelenti, hogy bemeneti adatok közötti hasonlóságból nem tudunk a képződő fehérjekomplexek eloszlásai közötti hasonlóságokra következtetni. Más szavakkal, különböző kiindulási gyakoriságok képesek hasonló jellegű komplexeloszlást eredményezni, és viszont, hasonló fehérjegyakoriságok nem eredményeznek feltétlenül hasonló komplexeloszlásokat.

A másik megközelítés alapján a bemeneteket elhelyeztem egy több dimenziós euklideszi térben, ahol minden tengely adott fehérje mennyiségét reprezentálja (bemeneti adatok). A kimeneti eredményeket szintén el kellett helyeznem egy többdimenziós térben. Ezt az összes felmerülő komplex sorbarendezésével tettem meg. A sorbarendezés oly módon történt, hogy előbb egy listába felvettem az összes előforduló komplexet, amihez egy Python szkriptet írtam, amiben a komplexeket mint gráfokat tekintem. Ezután a fehérjék adott sorrendje szerint és az azokat tartalmazó fehérjeszám szerint sorbarendeztem őket. Azaz minden komplex kapott egy kódot, amely olyan számsorból állt, ahol az egyes számok az adott

fehérje darabszámát jelentik a komplexben. Ez a rendezési reláció természetesen nem teljes, előfordulhatnak össze nem hasonlítható komplexek, ha minden fehérjéből pont ugyanannyi található meg bennük, mégis más kötésekkel. Ezen komplexek az összehasonlítás során egymás mellé kerülnek, így könnyen észrevehetőek és a logikát a sorbarendezésben nem zavarják meg. A sorbarendezés után minden komplex kapott egy saját tengelyt, így egy sok dimenziós térben a szimulációkat a kimeneti eredmények alapján is el tudtam helyezni. Majd normált távolságokat számoltam ki a minden bemenet-bemenet és minden kimenet-kimenet vektor között. A normálásra az eltérő dimenziószám miatt volt szükség. A normált távolságok viszont már összehasonlíthatóak voltak, adott szimulációpár bemeneti és kimeneti távolságának különbsége pontosan megmutatta, hogy mely agyi régiók, azaz szimulációk relatív helyzete változott meg a bemeneti helyzetekhez képest (8.6. ábra). **Ezáltal ki is tudtam emelni azokat a régiókat, amelyekben a legtöbb hozzáadott információ jelenhetett meg. Ezek a régiók: H376.VIII.51_STC id:340, H376.VI.50_V1C id:286, H376.IX.51_MFC id:238.** Az ID azt mutatja meg, hogy hányadikként végeztem el a szimulációkat, egyes ábrákon van szerepe. Az alulvonás előtti kód az alanyt jelenti, amelyből származik az agyi szelet, míg az alulvonás utáni háromkarakteres kód a konkrét agyi régióra utal. STC:posterior (caudal) superior temporal cortex (area 22c), V1C:primary visual cortex (striate cortex, area V1/17), MFC: anterior (rostral) cingulate (medial prefrontal) cortex.

Miután kidolgoztam a különböző távolságmétrikákat, megállapítottam, mely régiókra vonatkozóan tartalmazhatnak a szimulációk többletinformációt a bemenetekhez képest, és összevettem az outputokat az inputokkal - azaz a bemeneti fehérjék abundanciájának arányait összevettem a kimeneti komplexek arányaival, vagy a bennük található létrejövő páronkénti fehérje-fehérje kölcsönhatások relatív gyakorisága.

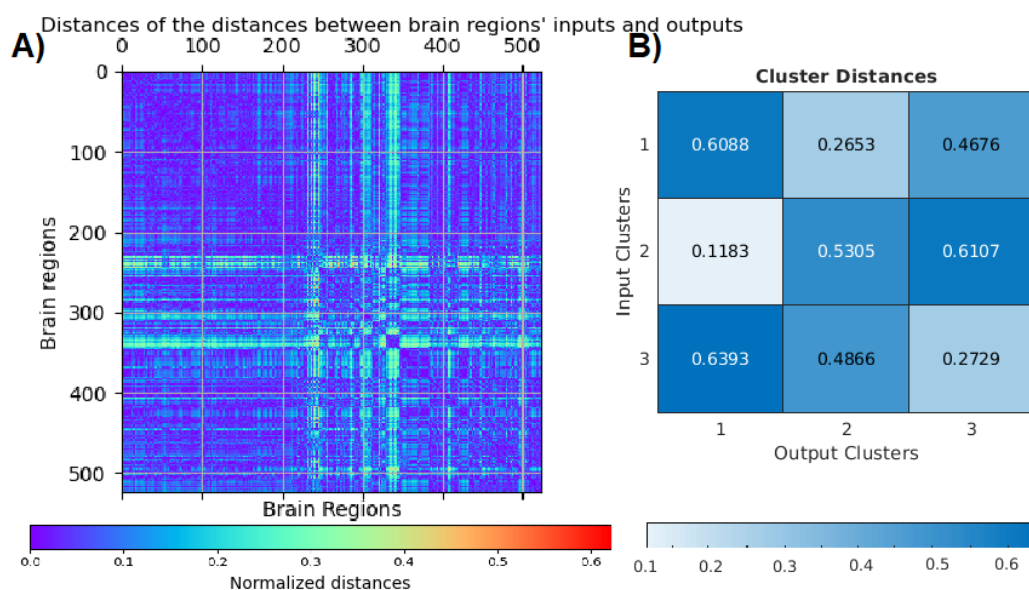
8.6.1 Az AMPAR/PSD-95/SYNGAP komplex a leginformatívabb

Főkomponens-elemzés segítségével azonosítottam, hogy mely komplexek a leginformatívab-
bak az agyi régiók megkülönböztetése szempontjából. Az első két fő komponens mind a vad típusú, mind a mutáns esetben a kimenetek teljes varianciájának 44%-át, illetve 24%-át fedi le. Az első főtengety az AMPAR/PSD-95 (id:12) komplex menyisége uralja, míg a másodikat a PSD-95/SynGAP (id:8) komplex.

Összességében a leginformatívabb komplex – az összes főkomponenst és az általuk magyarázott varianciát tekintve a legnagyobb hozzájárulással – az AMPAR/PSD-95/SynGAP (id:5) 19% hozzájárulással. Összehasonlításképpen, az átlagos komplex jelentőség nagyon kicsi, megközelítőleg $4,48e-06$, a medián pedig $9,98e-08$. Az alacsony mérték oka a lehetséges komplexek nagy száma - így a valószínűségi mérték nagy elemi eseményhalmazon osztozik eredendően is. Az, hogy a AMPAR/PSD-95/SynGAP komplex jelentősen befolyásolja a régiókat összhangban van az irodalomban fellelhető információkkal, SynGAP kötése regulálja az AMPAR kötésének gyakoriságát [168].

8.6.2 A régiótípusok számottevően nem különültek el egymástól

A főkomponens analízis szerencsére megerősítette azt a feltételezésemet, hogy a vizsgált régiók bemeneti és kimeneti eloszlása különbözik, azaz a szimulációk érdemi információval szolgálnak. Összességében a K-means algoritmus és a főkomponens analízis kiegészítheti egymást, és főtengetytranszformáció során is megfigyelhetőek a be- és kimenetek közötti

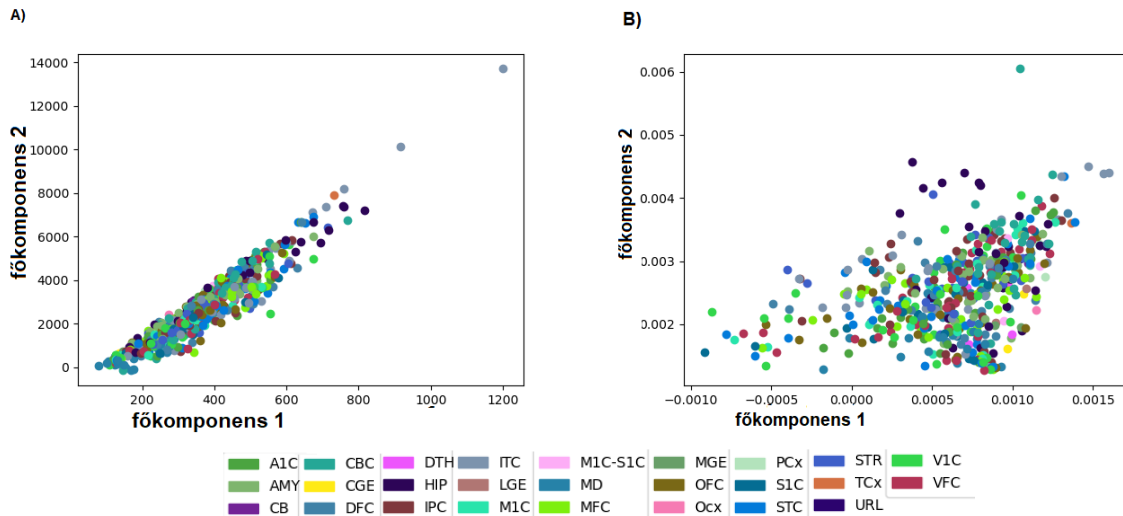


8.6. ábra. **A) A régiók egymáshoz képesti relatív elmozdulása hőterképben.** Az egyes régiókat mind bemenet, mind pedig kimenet alapján elhelyeztem egy adott vektortérben. Így a régiók relatív helyzete megmutatja, hogy két régió mennyire hasonlít egymáshoz, bemenet esetén a fehérjék mennyiségét tekintve, kimenet esetén az adott komplexek mennyiségét tekintve. Elmozdulás alatt azt értjük, hogy két régió a bemeneti helyzetükhöz képest közelebb vagy távolabb került egymástól a kimenet alapján megalkotott vektortérben. Sok esetben alig láthatunk változást, mert a legtöbb rész kék, mindazonáltal akadnak a pirosba is eső pixelek. **B) a klaszterezés eredménye.** A klaszterezés során nem egyértelmű, hogy melyik klasztert melyiknek érdemes megfeleltetni, ezért a legközelebbi klasztereket feleltettem meg egymásnak és azok között kerestem a megváltozás arányát. Az ábra a szerző saját munkája.

különbségek – többek között más a leginkább különálló régió is.

Azt is megvizsgáltam, hogy az anatómiailag egybeeső régiók által reprezentált adatsorok hasonlóan viselkednek-e. Ezt a főkomponens analízis nem támasztotta alá. Például az anterior medial prefrontal cortex régiók a szimuláció során nem viselkedtek hasonlóan és nem kerültek közelebb egymáshoz a kimenetben, mint amennyire a bemenetben voltak. Ugyanakkor előfordultak olyan agyterületek, amelyek jobban kivehetővé váltak az analízis után, a 8.7. ábra alapján. Így a másik, főtengeletytranszformációval kapcsolatos kérdésre is választ kaptam, miszerint annak során megjelentek olyan régiók, amelyek jelentősen magányosan találhatók meg, azaz nagyobb sugarú környezetükben sem találom más régiókat. Az ilyen, magányos régiók jellemzően különböznek a bemenetben tapasztaltaktól. Például a legmagányosabb, azaz a többitől legtávolabb eső régió a bemenetekben egy ITC régió, míg a kimenetekben a távolos ITC régiók továbbra is kicsit távolabb helyezkednek el a többitől, de már szorosabb csoportot alkotva. A kimenetekben a legmagányosabb régió egy CBC régió. A kimenetekben számos távolabb eső régiót is találunk. Megfigyelhető, hogy a kimenet szórása összességében véve nagyobb - azaz az átlagtól való átlagos eltérés, ami magának a szimulációegyüttesnek a normája, míg a bemenetekben a maximális távolság a számottevően

kiugró. Azonban az adatok ezen mutatóit nehéz összehasonlítani a kimenet és bemenet között, mert a tengelyek finomsága is más és más jelentőséggel is bírnak, tekintve, hogy más bázisban vannak felírva a dimenziószámok jelentős különbsége miatt. Összességében tehát az azonos agyterületi típusokat nem tudtam egyértelműen elkülöníteni. Ez adódhatott akár a bevezetésben említett neurontípusok folytonos varianciája miatt is, valamint amiatt, hogy egy agyi régió nem csak egy neurontípust tartalmazhat. Mindemellett azt is figyelembe kell vennünk, hogy az adatkészletben szerepelhettek kóros elváltozásokat mutató régiók is, melyek ezt az elemzést nagy mértékben befolyásolhatják, ezekről azonban nem találtunk információt.



8.7. ábra. **Főkomponens Analízis a A) bemeneti és B) kimeneti adatokon.** Az ábra a szerző saját munkája.

8.6.3 A komplexek eloszlásának alapvető jellemzőit elsősorban a fehérjegyakoriság határozza meg

Az elsőként elvégzett, a vad típusú esetet reprezentáló szimulációk során a hozzáadott információ tartalmát kerestem a konkrét eredmények között. Más szavakkal, az eredmények közül mi az, amit nem lehetett volna könnyen az eredeti fehérjeabundanciákból közvetlenül megjósolni. Megfigyeltem, hogy ez a hozzáadott információ tartalom valójában a fehérjék szomszédsági viszonyainak volt köszönhető. Adott szimulációk esetén a ténylegesen megjelenő komplexek gyakorisága eltért a hasonló fehérjegyakoriságok esetén kapottak alapján vártaktól. Az ilyen eltérésekre magyarázatot csak úgy kaptam, hogy összevettem, hogy adott fehérje mely másik fehérjével léphetett kapcsolatba, és az a másik fehérje milyen gyakorisággal szerepelt a rendszerben. **Arra jutottam, hogy a kölcsönhatásba lépő fehérjék abundanciája befolyásolja, hogy mely a fehérjét tartalmazó komplex jelenik meg. Felmerült a kérdés, hogy a kapcsolatokat befolyásoló mutáció jelenléte esetén ez megváltozik-e.**

A hipomorf Shank1 mutáció elemzésére futtatott szimulációk esetében ugyanolyan kiindulási fehérjegyakoriságokat használva futtatunk szimulációkat a vad típusú és a mutáns esetre is. A mutáns esetben homozigóta állapotot feltételeztem, tehát vagy csak vad típusú,

vagy csak mutáns fehérje volt a rendszerben.

Megfontoltam a heterozigóta eset szimulációját is, ami a számítások szintjén könnyen kivitelezhető, azonban az eredmények összehasonlításakor komoly nehézségeket okoz. Ennek oka, hogy a vad típusú és a mutáns Shank csak két különböző fehérjeként reprezentálható, és emiatt a lehetséges komplexek változatossága tovább nő. A változatosságnál azonban nagyobb probléma a kétféle Shank1 fehérjét tartalmazó komplexek összevetése a csak egyfélével (vad típusú vagy mutáns) tartalmazóakkal. Ezt a korábban kifejtettek értelmében a Cytocast beállítás szintjén nem lehet kezelni, mindenképpen a szimulációk végén kapott komplexek teljes újraannotálását kellene elvégezni, ami nagyon időigényes feladat, ráadásul helyessége is nehezen ellenőrizhető. Ezen ok miatt első közelítésben a heterozigóta eset szimulációjától eltekintettem.

Nem csak a bemeneti fehérjeabundanciák összehasonlításával érhetünk el eredményeket, hanem az is fontos, hogy a vad típusú és a mutáns típusú kimenetek közötti különbségeket is megvizsgáljuk. Ebben az esetben a mutáció hatásának mértékét régióspecifikusan értékelhetjük.

A vad típusú és mutáns forgatókönyvek szimulációs eredményeinek főkomponens-analízise nagyon hasonló összképet mutat. A két PCA diagram közvetlenül összehasonlítható, mivel a tengelyek még a két független PCA kimenetben is megegyeznek. A vad típusú és mutáns eseteknek megfelelő adatpontok csak minimálisan térnek el egymáshoz képest. A vad típusú és a mutáns régiók közötti átlagos távolság $1,5 \pm 0,8$ egység.

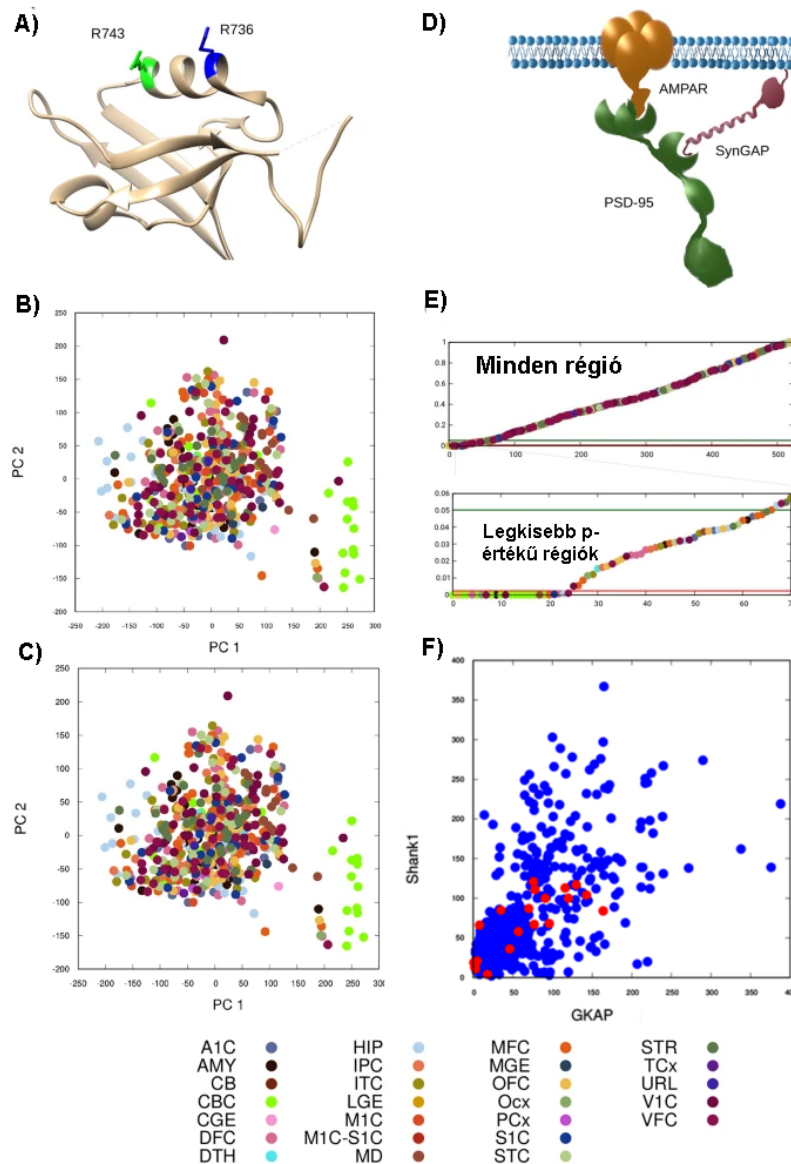
Megjegyzem, hogy a PCA általában nem különíti el az adott agyi régiókat (ábra 8.8), ezen szimulációk esetében azonban a cerebelláris kéreg típusú régiók azonban jól elkülönülnek a többtől.

Eredményeim arra utalnak, hogy a teljes fehérje komplex eloszlást elsősorban az egyes fehérjék elérhetősége határozza meg, és a vizsgált hipomorf mutáció jelenléte nem okoz jelentős globális hatásokat. Ez összhangban van azzal, hogy a rendszer megtartja általános funkcionalitását. A mutáció hatásának részletesebb elemzése érdekében megvizsgáltuk az egyes fehérjekomplexek abundanciáját.

Fontos megjegyezni, hogy **a fehérjekoncentráció szignifikánsan változhat az egyes alanyok között, ami a különbségek legnagyobb részéért felelős**. Amikor az agyi régiókra utalunk, figyelniük kell erre a jelenségre. Ezért különösen fontos, hogy ne tévesszen meg bennünket az egyének eltérő mintázata. Fontos megértenünk, hogy az ilyen eltérések nem feltétlenül jelentenek hibás eredményeket kutatásunkban, mivel az alapfeltevésünk az, hogy minden ember agya egyedi módon működik, és ez különösen kiemelkedhet az autizmus spektrum zavar esetében - ellenben oda kell rá figyelni.

Rábukkantam arra, hogy az adataimban a fehérjekoncentrációk alanyonként eltolódtak, amelyek befolyásolhatták az eredményeinket. Alanyok közötti kiugrások megfigyelése során például a PSD-95 mennyisége egy adott alanynál jelentősen több volt, mint a többinél - azt kerestem, hogy ez milyen változásokat okozhatott a szimulációkban és ezen változások mennyire jelentősek a többi szimuláció eredményéhez képest, azaz már a bemeneti különbség miatt lesz-e lényegesen különböző az eredmény is.

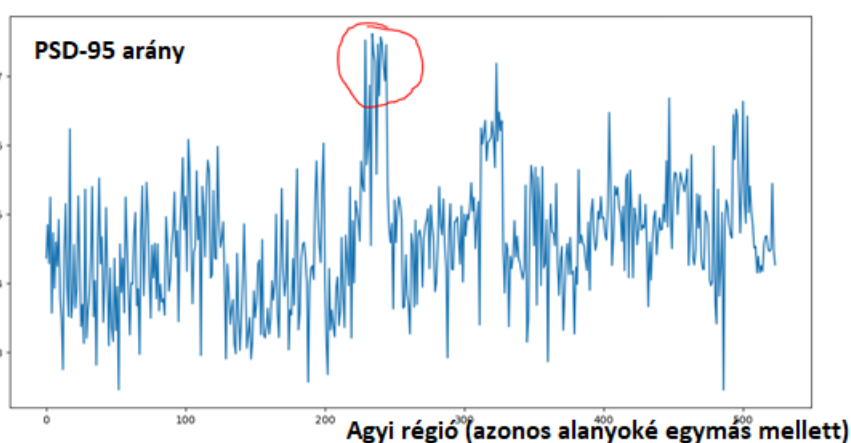
Tekintve, hogy a főtengelytranszformáció során nem, vagy csak korlátozottan tudtam azonos típusú agyi régiókat egyértelműen elkülöníteni a többi régiótól, ezért úgy határoztam, hogy számos újabb szimulációt futtatok le annak érdekében, hogy megvizsgáljam, hogy az azonos típusú agyi régiók miben térnek el egymástól. Ehhez például az összes anterior medial prefrontal cortex régiót megvizsgáltam külön-külön és egymással összehasonlítva is. a



8.8. ábra. **Mutáció hatása a szimulációra nézve:** **A)** A kiválasztott mutáció (R743H, zöld) és egy hasonló, az ASD-ben jelentett mutáció (R736Q, kék) pozíciója a Shank1 PDZ tartomány szalagábrázolásán (PDB ID 6YWZ). Mindkét arginin a ligandumkötő barázdát szegélyező $\alpha 2$ hélixen található. A kapott fehérje komplex eloszlások főkomponens-analízise a **B)** vad-típus és **C)** mutáns szcenáriók esetében. A különböző színek a különböző agyi régiókat jelölik az alján található kulcsnak megfelelően. **D)** A PCA szerinti leginformatívabb komplex sematikus ábrázolása (AMPA/PSD-95/SynGAP). **E)** p -értékek, amelyek leírják a mutáció változását a vad típushoz képest, az összes bemutatott komplexre átlagolt érték a leginformatívabb komplexre kapott értékeihez képest. A piros vonal az 0,01-es korrigált szignifikanciaszintet, a zöld vonal pedig az 0,05-ös szignifikanciaszintet jelzi. Ez segít azonosítani azokat a régiókat, ahol a mutációk hatása statisztikailag szignifikáns. Színek agyi régiók szerint az alján található kulcs szerint. **F)** A mutáció által érintett két fehérje, a Shank1 és a GKAP bősége a bemeneti adatkészletekben. Piros körök jelzik azokat az adatsorokat, ahol a leginformatívabb komplex bősége jelentősen megváltozott a kimenetben. Az ábra azt mutatja, hogy a két fehérje bősége és a szignifikancia között van valamilyen kapcsolat. Modus Tollens következtetés levonható, ha az abundanciák meghaladják a 150-t akkor a mutáció hatása elenyészik – tehát több fehérje a hatást csökkenti. Az ábra a szerző saját munkája, publikálásra került cikkemben [169].

legnagyobb változás ebben a régióban figyelhető meg és vannak olyan régiók, melytől még kiemelkedőbben eltér, annak ellenére, hogy mindegyik régió az anterior prefrontal cortex agyterületnek felel meg. Ez arra engedett következtetni, hogy adott régiótípuson belül is lehetnek jelentősebb különbségek az elsődleges komplexek arányában.

Ebben is kiemelkedett a 238-as anterior medial prefrontal cortex régió, mely bemeneti adataira jellemző, hogy jelentősen kiemelkedik a PSD-95 fehérje gyakorisága. A kiemelkedő PSD-95 gyakoriság az alany majdnem összes régiójára jellemző (lásd 8.9. ábra). Az ábrán adott alany összes agyi régiója egymás mellett található. Ennek a kiemelkedésnek az okára a források kettős ellenőrzése során sem kaptam magyarázatot, de ebben az esetben nem kizárható sem kísérleti hiba, sem valamilyen elváltozás megléte. Az alany vizsgálatainkból való kizárására nem láttunk okot, hiszen a nagy mennyiségű bemeneti adatra az előforduló változatosság maximalizálása miatt volt szükségünk.



8.9. ábra. **A PSD-95 fehérje aránya minden agyi régióban** Az ábrán mind az 524 agyi régió látható. Azonos alanytól származó régiók egymás mellett helyezkednek el. Az ábra a szerző saját munkája. A piros karika jelöli azt az alanyt, melynek legtöbb agyi régiójában a PSD-95 gyakoribb, a többi alany régiójához képest.

8.7 Különböző fehérjekapcsolatok hatásai a komplexekre

8.7.1 Homer1 tetramerizáció és Shank3 szelfasszociáció hatásai

Szimulációimat három változatban futtattam, a legrealisztikusabb, a **H4SM** elnevezésű – ide tartoznak az eddig bemutatott eredményeim – figyelembe veszi a Homer1 tetramerizációt a C-terminális coiled coil régióján keresztül és a Shank3 multimerizációt a SAM doménjén keresztül. A **H4** jelű szimulációk csak az előbbit tartalmazzák. Ezekben felül futtattam egy olyan változatot, ahol ezen homomultimerizációs kölcsönhatások egyike sem szerepelt. A Homer1 dimerek dimerjét képezheti [170], és az egyszerű szimulációban csak azokat a dimereket vettem figyelembe, amelyek még mindig képesek bivalens kölcsönhatások kialakítására az egyes monomerek EVH1 doménjén keresztül.

A szimulációk tényleges kérdése viszont az, hogy adott szimulációban milyen komplexek jelennek meg és azok milyen mennyiségben jelennek meg. Két szimulációt nehéz összehasonlítani, ha konkrét darabszámokat tekintünk, hiszen a kiindulási komponenseink, a

fehérjék is különböző abszolút mennyiségben lehetnek jelen. Így az összehasonlításhoz a relatív gyakoriságokat vettem figyelembe, azaz kördiagramon ábrázoltam, hogy az adott szimulációban előforduló komplexek hányad részét teszi ki az adott komplex. Így meg tudtam határozni, hogy mely komplexek a leggyakoribbak a szimuláció szerint adott agyi régióban.

A legösszetettebb rendszer, amit vizsgáltam, figyelembe vette a Homer1 tetramerizációját és a Shank3 szelfasszociációját is. Ez az elsődleges komplexek szimulációjához képest egy bővített rendszernek számít. Megvizsgáltam, hogy a Homer1 tetramerizációja hogyan hat a rendszerre, és kiértékeltem a szimulációkat, melyben már az elsődleges komplexekhez képest jóval összetettebb komplexeket is kaptam eredményül a vártak megfelelően.

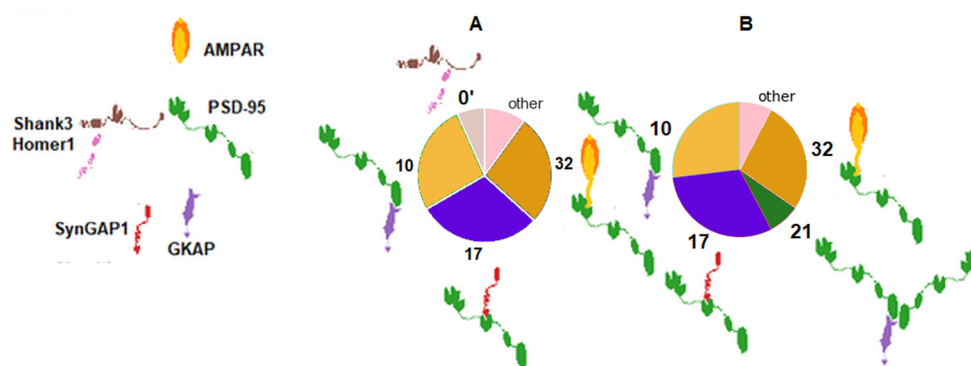
A szimulációban többféleképpen is meg lehet valósítani egy tetramerizációt. Azt a megoldást választottam, melyben a rendszerbe a Homer1 molekulákat már eleve dimerizált formában adtam meg, azaz a molekulát reprezentáló ágensen kétszer annyi kötőhely volt, mint egy eredeti monomer Homer1 molekulán lett volna. A tetramerizáció pedig így két már dimerizált Homer1 dimerizációjaként valósult meg a szimuláció során. Természetesen emiatt a Homer1 gyakoriságát a monomerre kapott érték felére kellett csökkenteni és egész számra kerekíteni. Az, hogy ez a módszer mennyit ront a pontosságon, attól függ, hogy természetes állapotában inkább monomer, vagy inkább dimer formájában található-e meg jobban a fehérje. Az utóbbi esetben a módszerem pontosabb, míg az első esetben érdemesebb lehet szétszedni a dimereket és a tetramerizációt kötésről kötésre létrehozni - ezzel a tetramerek gyakorisága csökken. Tekintve, hogy a Homer1 dimerizáció coiled coil szerkezeti motívumon keresztül valósul meg, biológiailag releváns az obligát dimerként való modellezése.

A Homer1 tetramerizációt tartalmazó szimulációk hasonló relatív fehérjegyakoriságokat eredményeztek, mint az elsődleges komplexek - azaz amilyenek a tetramerizáció nélküli szimulációk során előfordultak. Ugyanazok az elsődleges komplexek rendelkeztek a legmagasabb gyakorisággal, azonban több „egyéb” kategóriába sorolt komplex fordult elő amelyeknek a gyakorisága kisebb volt, mint az ábrázoláshoz meghatározott küszöb – az ez alatti példányszámban megjelenő komplexeket egységesen az „egyéb” kategóriába soroltam a kördiagramokon. Ez azt mutatta, hogy több olyan komplex is megjelent, ami eredetileg tetramerizáció nélkül nem létezhetett. Ezek az összetettebb komplexek voltak azok a komplexek, amelyek Homer1 tetramerizációjának segítségével tudtak létrejönni. Önmagában véve a Homer1 tetramerizációja nem hozta meg a várt szuperkomplexek létrejöttét. Csupán két teljes vagy hiányos komplex-együttes csatlakozott az egyébként négy lehetőségből a tetramerekre. Ez azt mutatja, hogy noha négyszer akkorra komplex létrejöttére van lehetőség a tetramerizáció során, ez az egyes fehérje-fehérje kapcsolatok létrejöttének valószínűsége miatt mégsem feltétlenül valósul meg, legalábbis a szimuláció körülményei között.

Tekintve, hogy a Homer1 tetramerizációja nem adott választ a szuper-komplexek létrejöttére, ezért a legösszetettebb rendszert is megvizsgáltam. Ez alatt azt a szimulációt értem, amelyben a Homer1 tetramerizációján túl a Shank3 fehérjék képesek az önmagukkal való asszociációra. A SAM doménjükön keresztül a Shank3 fehérjék önmagukban is képesek egymáshoz kötődni, és azonos fehérjékből álló hosszú láncokat kialakítani. Ezek a nagyméretű láncok így lehetőséget biztosítanak arra, hogy minden egyes Shank3-hoz tudjon kötődni egy fehérjekomplexekből álló oldalág (későbbiekben erre komplexláncként fogok hivatkozni) és egy Homer1 tetramer, ami további három láncnak ad lehetőséget csatlakozni. Minden ilyen, a tetramerhez csatlakozó további láncon a Shank3 újabb Shank3-láncokat képes létrehozni, ezáltal növelve az újabb komplex-láncok kialakulásának lehetőségét. Ez a mechanizmus

szimulációim során meg is történt. Különösen érdekes eredménynek tartom, hogy a plusz kötési lehetőségek az eredeti eredményeken az elsődleges komplexek arányát tekintve nem vagy minimálisan változtattak. Ez azt jelenti, hogy már a Homer1 tetramerizáció és Shank3 önasszociáció nélkül is jól megjósolható, mely komplextípusok kerülnek túlsúlyba - melyek lesznek az alapvető építőelemei a bonyolultabb komplexeknek. Természetesen az arányok megmaradása a jóval nagyobb, azaz több fehérjét tartalmazó komplexek esetében már nem mondható el, hiszen több száz fehérjét tartalmazó szuperkomplexek a Shank3 önasszociáció lehetősége nélkül nem jöttek létre. Az elsődleges komplexekhez képest jóval kevesebb, már "elhanyagolható" mennyiségben jelentek meg a nagy szuperkomplexek. Mindazonáltal a fehérjekomponensek számára vonatkozó egy adott küszöbindexnél nagyobb komplexek esetében már a kevés 1-2 darab komplex is jóval nagyobb befolyással van a poszt-szinaptikus denzitásra, hiszen méretéből adódóan egy szuperkomplex kitöltheti majdnem akár a teljes denzitás java részét is. Ebből adódóan adott méret után már nem összehasonlítható a kisebb és elsődleges komplexek gyakoriságával a nagyobb szuperkomplexek jelentősége. Erre dinamikus szignifikancia határt is ki lehet dolgozni, ami a komplex méretétől (esetleg a térbeli szerkezetétől - a térszerkezetek modellezésének egy visszacsatolási lehetősége a szimulációkra) függ.

A legnagyobb komplex 75 komplexláncot kötött össze egymással és 4240 fehérjét tartalmazott. A szimulációk szerint a Shank3 fehérje felelős a legtöbb receptor egy komplexben való megjelenéséért. A legnagyobb komplex 49 receptort kapcsolt össze. Ebből kifolyólag a modellünk szerint a Shank3 génjeinek kiütése vagy expressziójának csökkentése széteső szuperkomplexeket eredményez.



8.10. ábra. **A Homer tetramerizáció önálló hatásának vizsgálata:**A 238-as régió eredményei A) Egyszerű beállítás B) H4 beállítás. A fő komplexek ugyanazok maradtak, mint a H4SM beállításnál. Az ábra a szerző saját munkája.

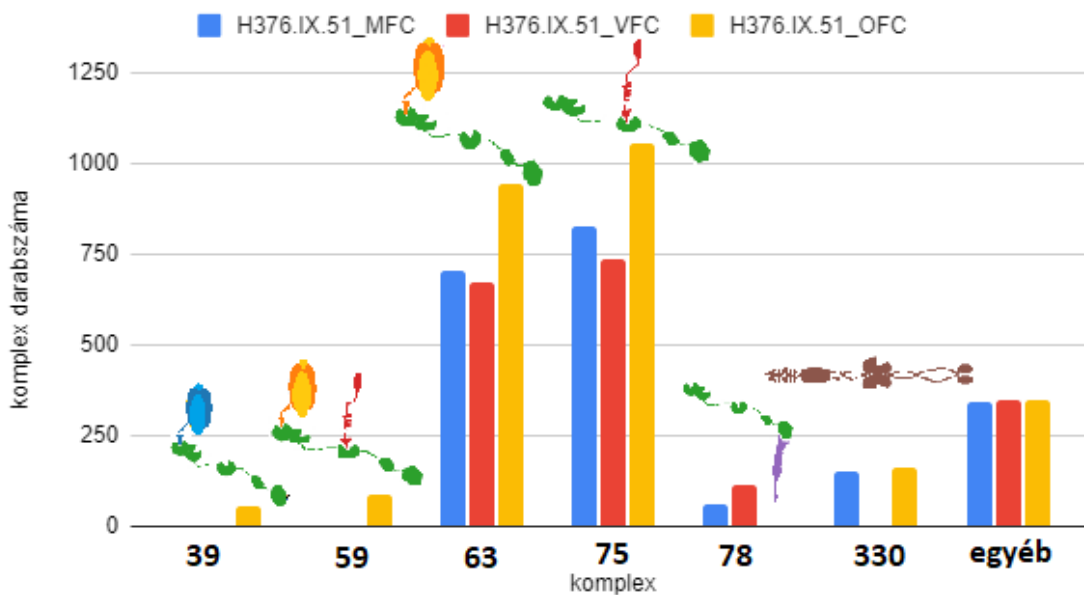
Összefoglalva, az egyszerű szimulációimban az elsődleges komplexek relatív abundanciája hasonló volt a H4 és H4SM szimulációkhoz (8.10. ábra), de nagy szuperkomplexek nem jöttek létre. Általánosságban elmondható, hogy a Shank3 multimerizáció és a Homer1 tetramerizáció növelte a kapott komplexek diverzitását, túlnyomórészt az elsődleges komplexek asszociációján keresztül. A Homer1 tetramerizáció megengedésével önmagában még nem, de a Shank3 önasszociáció mellett már megjelennek szuperkomplexek.

8.7.2 A hasonló fehérjegyakorítások ritkán, de eredményezhetnek eltérő komplexeloszlást

A szimuláció információtartalmát vizsgálva három régiót találtam kiemelendőnek, a régiók: H376.VIII.51_STC id:340, H376.VI.50_V1C id:286, H376.IX.51_MFC id:238. Így ezeket a régiókat részletes elemzésnek vettem alá. Mindegyikhez megkerestem a bemenet alapján két-két legközelebbi régiót és megvizsgáltam az eltéréseket a szimulációk kimenetében. Az elsődleges komplexek arányát tekintve bizonyos fokig hasonló eredményeket kaptam, azonban minimális eltérések észrevehetőek voltak. Ezek már biztosan nem a szimuláció sztochasztikus jellegéből adódhattak, mert azok hatását egy átlagoló eljárással nagymértékben redukáltam.

A vizsgált régiók közül a H376.VI.50_V1C és a H376.VIII.51_STC (283 és 340) egymáshoz közeli régióknak számít a kiszámolt metrika alapján, ezáltal hasonló régiókként érdemes elemezni az eredményeiket, mégha különböző agyi területeknek is felelnek meg. Ezek esetében valóban nagyon hasonlítottak a fehérjegyakorítások, mindössze az NMDA receptorok és a Homer1 fehérjék száma volt kicsit eltérő egymáshoz képest. Ahogyan vártam, a szimulációk során az elsődleges komplexek közel azonosak voltak, és majdnem azonos mennyiségben is jelentek meg. Akadtak azonban kisebb eltérések, mert például az utóbbi régióban nem jelent meg számottevő mennyiségben a Shank3 dimer, mely arra enged következtetni, hogy a Homer1 aránya a Shank3-hoz képest képes eltolni a Shank3-Shank3 kötések kialakulásának valószínűségét is. Ezzel együtt még a PSD-95/SynGap1 komplex megjelenésében is változásokat figyeltem meg.

Három hasonló régió összehasonlítása



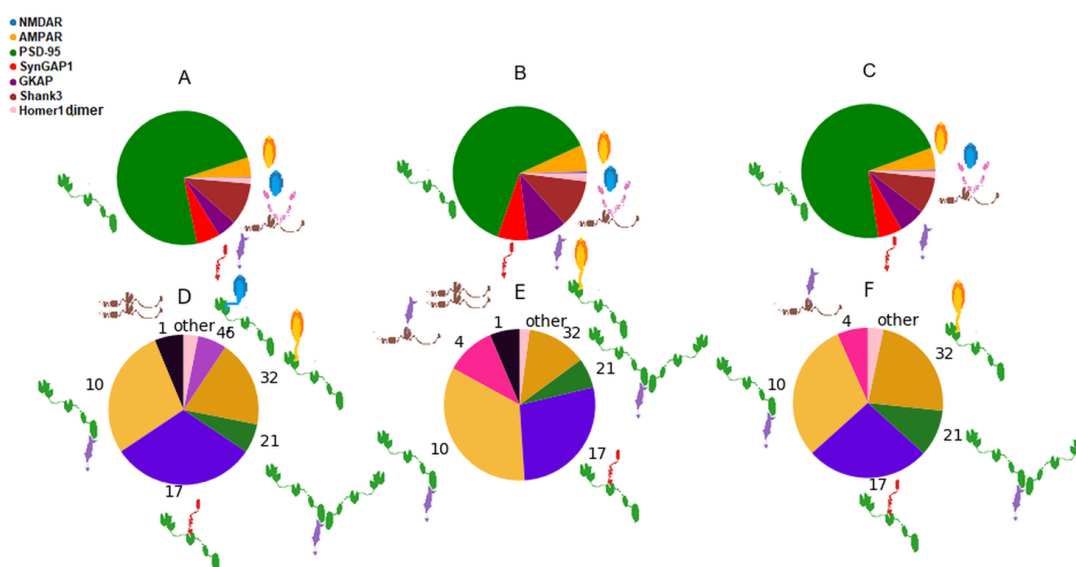
8.11. ábra. A legkiugróbb MFC régióhoz a két leghasonlóbb bemenetű régió szimulációinak összehasonlítása. Az ábra a szerző saját munkája.

238. régió: elülső mediális prefrontális kéreg nagy mennyiségű PSD-95-tel

Az alfejezet eredményeit a [171] cikkben közöltem.

A legnagyobb különbséget a kimenetben a bemenethez képest a 238-as agyi régióban (H376.IX.51_MFC) figyeltem meg, amely az elülső mediális prefrontális kéreghez tartozó adatokat reprezentál. Ebben a régióban a PSD-95 az összes fehérje több mint 75%-át teszi ki, ami szokatlanul magas, mivel a többi régióban csak ritkán haladja meg az 55%-ot. A PSD-95 óriási aránya alanspecifikus. A H376.IX.51 alany minden régiójában hasonló mennyiségű PSD-95-öt tartalmaz, körülbelül 75%, de az adatforrásból nem derül ki, hogy diagnosztizáltak-e valamilyen idegrendszeri betegséget ennél a személynél ([30]).

A receptorokat tekintve az AMPAR abundanciája jóval magasabb, mint az NMDAR-é, amelynek mennyisége elhanyagolható. A Shank3 a második legnagyobb mennyiségben előforduló fehérje, míg a GKAP csak kis mennyiségben van jelen. A kimeneti komplexeket a PSD-95/GKAP, PSD-95/SYNGAP és AMPAR/PSD-95 bináris kölcsönhatások uralják. A Shank3 főként a Shank3 dimerekben található (lásd 8.11. ábra).



8.12. ábra. **A bemeneti fehérje és a kimeneti komplex mennyiségének összehasonlítása a 238-as régióban és a két leginkább hasonló bemeneti fehérjeadatokkal rendelkező régióban** A, B és C: Bemeneti fehérje mennyisége a 238-as, 239-es és 244-es régióban. D, E és F: a 238-as, 239-es és 244-es régió kimeneti komplexeinek mennyisége. A kördiagramok a bemeneti fehérjék/kimeneti komplexek frakcióit mutatják. A 2-nél kisebb abundanciájú komplexeket összevonjuk, és „egyéb” komplexként jelenítjük meg. A diagramok azt mutatják, hogy a bemeneti fehérje bőség arányának viszonylag kis változása a komplex frakciók jelentős újraeloszlásához vezethet. A legnagyobb változást mutató komplexek közül kettő (Shank3 dimer, id 1; és Shank3/GKAP, id 4) nem tartalmazza a legnagyobb mennyiségben előforduló PSD-95 fehérjét, és megjelenésük nem is triviálisan függ alkotóelemük arányától. Az ábra a szerző saját munkája.

Azonosítottam a 238-as régióhoz leginkább hasonló inputokkal (fehérjegyakorisági adatokkal) rendelkező két régiót, és összehasonlítottam a megfelelő kimeneteiket (komplex

eloszlás) egymással. Bár a bemenetek nagyon hasonlóak (lásd a 8.12 ábrát), figyelemre méltó különbségek figyelhetők meg a kimenetekben, amelyek világos példát adnak a fehérje- és a komplexgyakoriság közötti nem triviális kapcsolatokra, és jól szemléltetik az alkalmazott szimulációk hozzáadott információtartalmát.

Mindhárom régió ugyanattól az egyéntől származik, egy minta a ventrolateralis prefrontális kéregből (244. régió), egy másik pedig az orbitális frontális kéregből (239. régió). Ezekben a régiókban a Homer1 és a SynGAP1 előfordulása mutatja a legnagyobb különbséget a 238-as régióhoz képest. A régiókhoz tartozó komplexeloszlások közötti legszembetűnőbb különbség azonban a Shank3 dimerek jelenlétében mutatkozik, amelyek hiányoznak a 244-es régióból. Ehelyett a Shank3 fehérjék több nagyobb asszociátumban találhatóak: komplexekben vagy a GKAP/Shank3 bináris komplexben. Jól látható, hogy a valamivel kevesebb GKAP nem eredményez magasabb Shank3 dimer arányt – amit egyébként várhatnánk a GKAP/Shank3 kölcsönhatások kialakulásának csökkent esélye miatt –, ami nem triviális kapcsolatot jelez a bemeneti fehérje gyakorisága és a komplexképződés között a szimulációinkban.

254-es és 339-es régió: hasonló bemeneti mennyiségek eltérő kimenetei

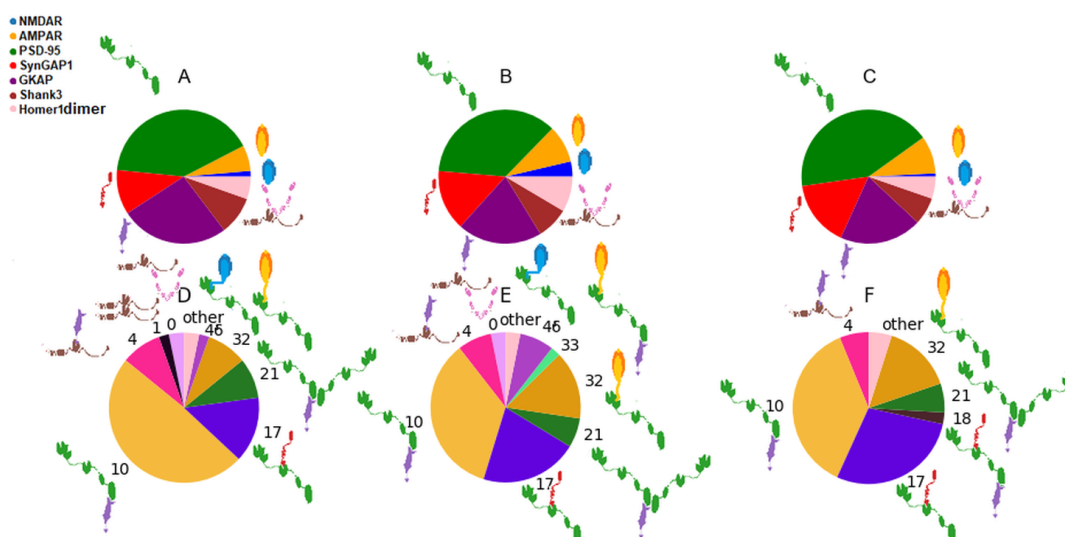
Az alfejezet eredményeit a [171] cikkben közöltem.

A 254. régió, amely megfelel az elsődleges látókéregnek (H376.IX.52_V1C) és a 339. régió, mely az elsődleges szenzoros kéregnek felel meg (H376.VIII.51_S1C), a második és a harmadik legnagyobb távolságkülönbséget mutatta a bemeneti és kimeneti adatvektorok között. Érdekes módon ennek a két bemeneti fehérjekészletnek mennyisége csak kismértékben különbözik. A két bemenet a SynGAP1, GKAP és AMPAR fehérjék arányában különbözik.

A 254-es és 339-es szimulációk komplexei közötti fő különbség a PSD-95/GKAP és a SynGAP1/PSD-95 arányban rejlik, amint azt a bemeneti különbségek alapján is várhatjuk. Azonban mindkét szimulációban a várakozásoknak megfelelően megjelenik a SynGAP1/PSD-95/GKAP hármas komplex, de kevesebb SynGAP1 és több GKAP nem vezet a PSD-95 hármas komplexének kialakításához. Ebben az esetben (254. kísérlet) a PSD-95(2)/GKAP magasabb aránya kizárja, hogy a GKAP csatlakozzon a SynGAP1/PSD-95 bináris komplexhez.

Mindkét régió esetében azonosítottam azokat a régiókat, amelyek a legnagyobb hasonlóságot mutatják a bemeneti mennyiségek tekintetében. A 254-es régió esetében (lásd a 8.13 ábrát, a legtöbb hasonló fehérjegyakoriságot mutató régiók komplex eloszlása is nagymértékben hasonlít egymáshoz. Számszerűen a legnagyobb eltérések a 10-es komplexeknél figyelhetők meg (PSD-95/GKAP) A kördiagramokon az 1-es komplex (Shank3(2)) látszólag hiányzik a másik két régióból, és a 18-as hármas komplex (SynGAP1/PSD-95/GKAP) csak a 254-es régióban fordul elő. Ennek azonban csak annyi az oka, hogy gyakoriságuk éppen az ábrázolás során az egyedi címkézésre vonatkozó küszöbérték (2) alatt van.

A 339-es régió hasonló különbségeket mutat a két legközelebbi régióhoz képest a fehérjegyakoriság tekintetében, és a legnagyobb különbség a 17-es (SynGAP1/PSD-95) és a 10-es (PSD-95/GKAP) komplexek előfordulásában is a legnagyobb, ugyanakkor a legnagyobb arányú (8.14). Itt, bár a 32-es komplex (AMPAR/PSD95) abszolút mennyisége nagyon hasonló (13, 14 és 17 komplex a 339-es, 328-as és 340-es régióban), mégis ezek relatív aránya az összes komplexhez képest eltérő.



8.13. ábra. **Bemeneti fehérje és kimeneti komplex mennyiségek a 254. kísérlethez és a két legtöbb hasonló bemenetet tartalmazó régiókhoz.** A) A 254. régió bemeneti fehérje mennyisége, B) a 342. régió bemeneti fehérje mennyisége, C) a 255. régió bemeneti fehérje mennyisége, D) a 254. régió kimeneti komplex mennyisége, E) a 342. régió kimeneti komplex mennyisége, F) a 255. régió kimeneti komplex mennyisége. Az ábra a szerző saját munkája.

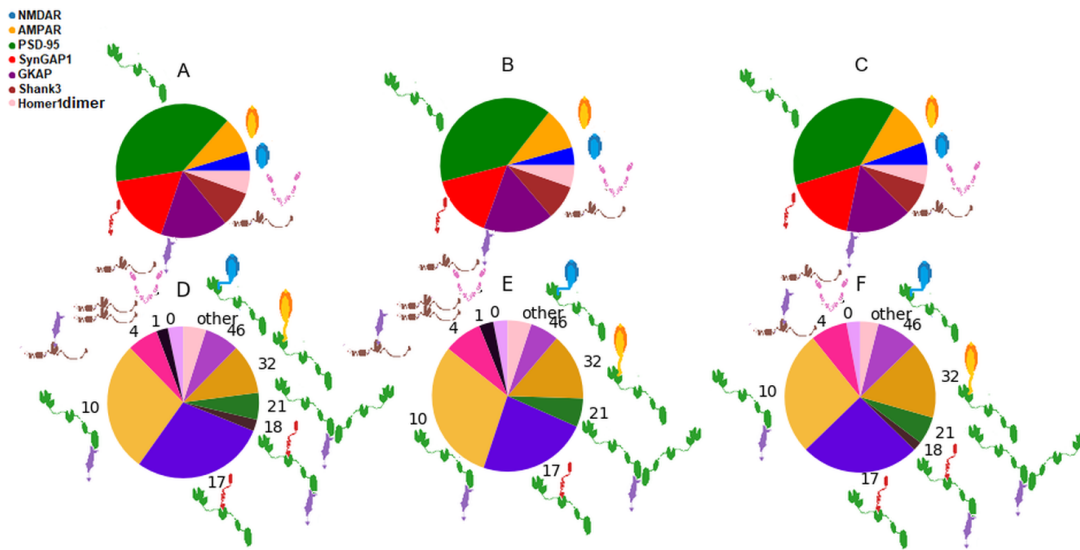
8.7.3 A Shank:GKAP interakció gyengítése korlátozott változásokat okoz egyes jól meghatározott régiókban

A vizsgált mutáció hatásának elemzéséhez összehasonlítottam az egyes régiók komplex eloszlását a vad típusú szimulációkkal. Először a kumulatív komplex eloszlásokat hasonlítottam össze az egyes komplexekre végzett *t*-próbák súlyozott átlagának segítségével, ahol a súlyok a komplexek információtartalmán alapultak (lásd Módszerek 8.5.3). Ezek a tesztek azt mutatják, hogy a kumulatív komplex eloszlások a statisztikai szignifikancia határain belül változatlanok maradnak - lásd a 8.8 ábrát - annak ellenére, hogy a Shank1-GKAP kölcsönhatás disszociációs sebessége ötször nagyobb a mutánsban. Bár a kumulatív *p*-értékek nagy eltérést mutatnak, mindegyik 0,05 felett marad, így általában a régiók és komplexeik többségénél a nullhipotézist 5% szignifikanciával fogadjuk el.

A nagyobb szuperkomplexek esetében kisebb gyakoriságváltozás is nagyobb szignifikanciával rendelkezik mint a kisebb komplexek esetében, mivel előbbiek sokkal kisebb valószínűséggel alakulnak ki. Ezért a szuperkomplexek esetében a legkisebb megjelenés is elfogadott és megfelelőnek tekinthető. A *p*-értékek varianciája azonban arra utal, hogy a mutáció nem egyenletesen érinti a régiókat, még akkor sem, ha az összesített változás nem éri el a statisztikai szignifikanciát.

Az egyes komplexek szintjén azonban néhány jelentős változás mutatható ki a mutációt hordozó esetben a vad típusúhoz képest.

Fontos, hogy a leginformatívabb komplex, az AMPAR/PSD-95/SynGAP -értékei 0,05 alá esnek az eredményül kapott adatkészletek némelyikénél (8.8 ábra). Ez némileg meglepő,

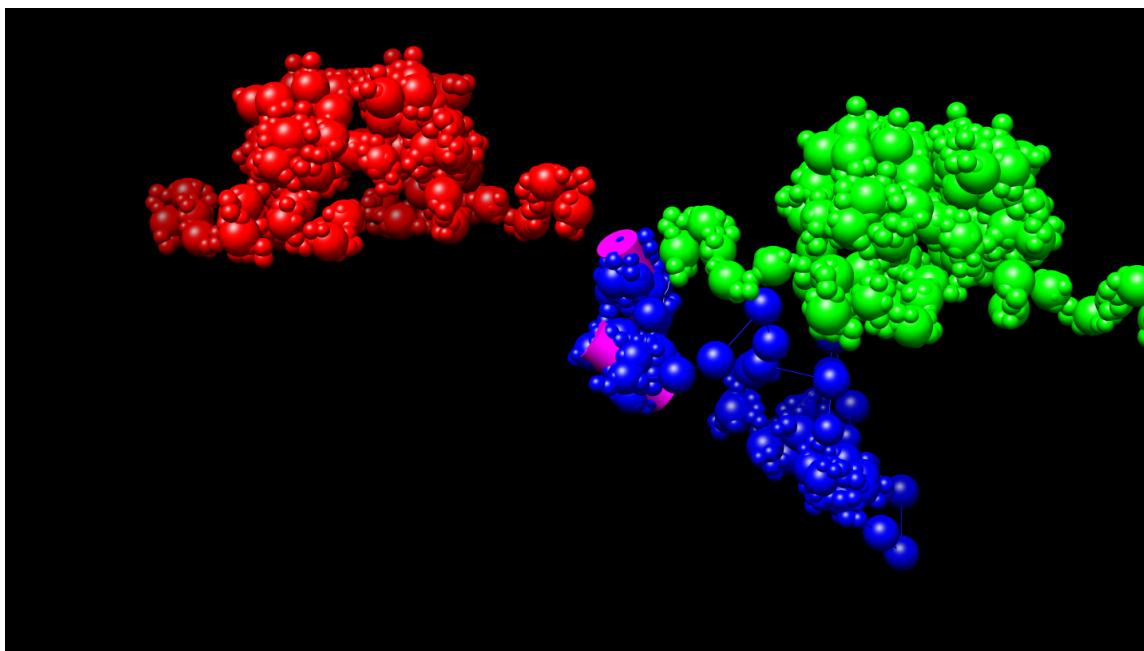


8.14. ábra. **Bemeneti fehérje és kimeneti komplex mennyiségek a 339. kísérlethez és a két legtöbb hasonló bemenettel rendelkező készlethez.** A) A 339. régió bemeneti fehérje mennyisége, B) a 328. régió bemeneti fehérje mennyisége, C) a régió bemeneti fehérje mennyisége 340, D) a 339. régió kimeneti komplex mennyiségei, E) a 328. régió kimeneti komplex mennyiségei, F) a 340. régió kimeneti komplex mennyiségei. Az ábra a szerző saját munkája.

mivel ez a komplex nem tartalmazza a mutáció által érintett kölcsönhatás egyik partnerét sem. Érdekes tehát megvizsgálni, hogy az egyes partnerfehérjék Shank1 és GKAP mennyisége hogyan befolyásolja az AMPAR/PSD-95/SynGAP komplex kialakulásában megfigyelt változásokat. Ábrázoltam a Shank1 és a GKAP bemeneti mennyiségét a leginformatívabb komplexre kapott p -értékekkel szemben (8.8 ábra). Nyilvánvaló, hogy a jelentős változások nem korlátozódnak a két érintett fehérjét különösen nagy vagy alacsony gyakorisággal tartalmazó régiókra, ami arra utal, hogy a megfigyelt jelenség háttérben több fehérje összetett kölcsönhatása áll.

A mutáció eredményeként azt várjuk, hogy a GKAP feletti rétegben lévő komplexek, azaz a membránreceptorokat és a PSD-95-öt tartalmazó komplexek jellemzőbbé váljanak, mivel a GKAP és a Shank1 közötti kölcsönhatás összekapcsolja ezeket a komplexeket a nagyobb szuperkomplexekkel, ahol a Shank1 polimerizálódik. Ez a hatás csak a cerebelláris kéreg régiókban és ott is csak kis mértékben figyelhető meg. A gyakoriságokat a S2 ábra mutatja. A komplexek mennyisége nulláról 1-re változik a SynGAP/PSD-95/GKAP (id:9) és AMPAR/PSD-95/GKAP (id:15) komplexeknél. A többihez képest azonban ezen komplexek mutáns sejtekben való megjelenésének jelentősége még mindig nagyon alacsony, ami azt jelzi, hogy a komplexek eloszlása főként a legkisebb p -értékű régiókban is hasonló marad a vad típusra jellemzőhöz, így a választott szignifikancia szint elfogadható.

Érdekes módon a legalacsonyabb p -értékkel rendelkező régiók mind a kisagykéreghez tartoznak. A kisagy szerkezete és szintjei szignifikánsan eltérnek a nagyagytól, ami a fő neurontípusok eltéréseit is jelenti [172]. A különbségeket már kimutatták a Shank3 és Shank2



8.15. ábra. A PDZ domének hengerként, ahol a káliumcsatornák közötti távolság $36,65\text{\AA}$. Az ábra a szerző saját munkája.

fehérjék előfordulásának esetében a kisagy különböző rétegeiben [173], és a kisagy különböző aspektusaiban az ASD-ben [174]. Közvetlenül nem állítható, hogy a p -értékekben megjelenő információ erre a különbségre utalna, de nem kerülheti el figyelmünket.

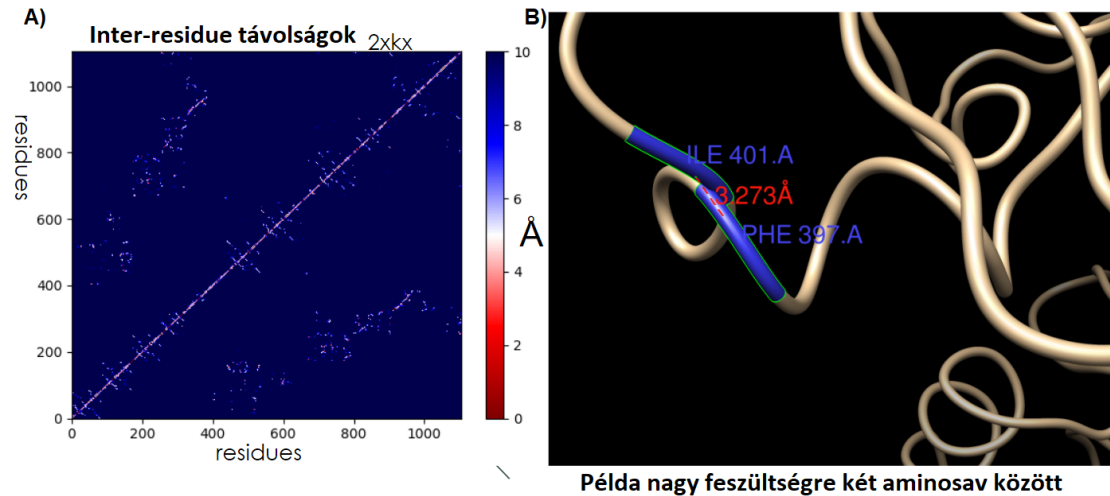
Ahhoz, hogy további betekintést nyerjünk a komplexek gyakoriságának változásai, két régiót választottunk ki: a H376.IIIB.53_M1C-S1C a legalacsonyabb nem nulla p -értéket mutatja a leginkább informatív komplex (AMPA/PSD-95/SynGAP) és a H376 esetében. A IV.54_STR p -értéke a 0,05-ös szignifikanciaküszöb alatt van ugyanarra a komplexre vonatkozóan (S3 ábra).

8.8 A felépített háromdimenziós modellek

Az első fehérje, amely esetében a teljes 3D szerkezeti modell összerakását megkezdtem, a PSD-95 volt. Az összerakás során azt tapasztaltam, hogy a fehérje terminális szakaszait így is nehéz megjeleníteni térszerkezet hiányában, mert a fehérjeszekvenciák alapján létrehozott kezdetleges szerkezeti modellek nem lettek megbízhatóak, mert számos helyen jelentősen eltértek a torziós szögek az elfogadott tartományoktól. Az IMP eljárás segítségével épített modellekből megállapítottam, hogy egy PSD-95 70-75Å között hidal át két Kir2.1 molekulát a PDZ tandem keresztül. A PDZ doméneket hengerként tekintve azt kapjuk, hogy a PDZ domain átlagos hossza 31Å és a sugara 7Å a 8.15. ábrán lilával látható. Ez lehet az egyik korlát, amiért a 35Å-nál magasabb távolságok esetében a káliumcsatornák között a PSD-95 ingázik [175].

A PSD-95 esetében egyetlen olyan szerkezetet találtam a PDB adatbázisban, amely a teljes fehérjét lefedte, ez a 2xkx PDB kódú szerkezet. Ezen a szerkezeten is végimentem az általam meghatározott validálási lépésekkel és arra jutottam, hogy a szerkezet nem annyira

megbízható. Megnéztem az aminosavak közötti sztérikus ütközések kiszámításával (lásd 8.16 ábra). Egy átlagos aminosav távolságot megnevezhetünk, mint küszöbindex, mert ebben az esetben nem aminosavanként vizsgáljuk a szerkezetet, de az aminosav szekvencia szerinti sugarakkal is végig lehet rajta menni és akkor egy részletesebb felbontást kapunk. Az AlphaFold szerkezetekkel még nem próbálkoztam, de azokkal tervezek továbbhaladni.



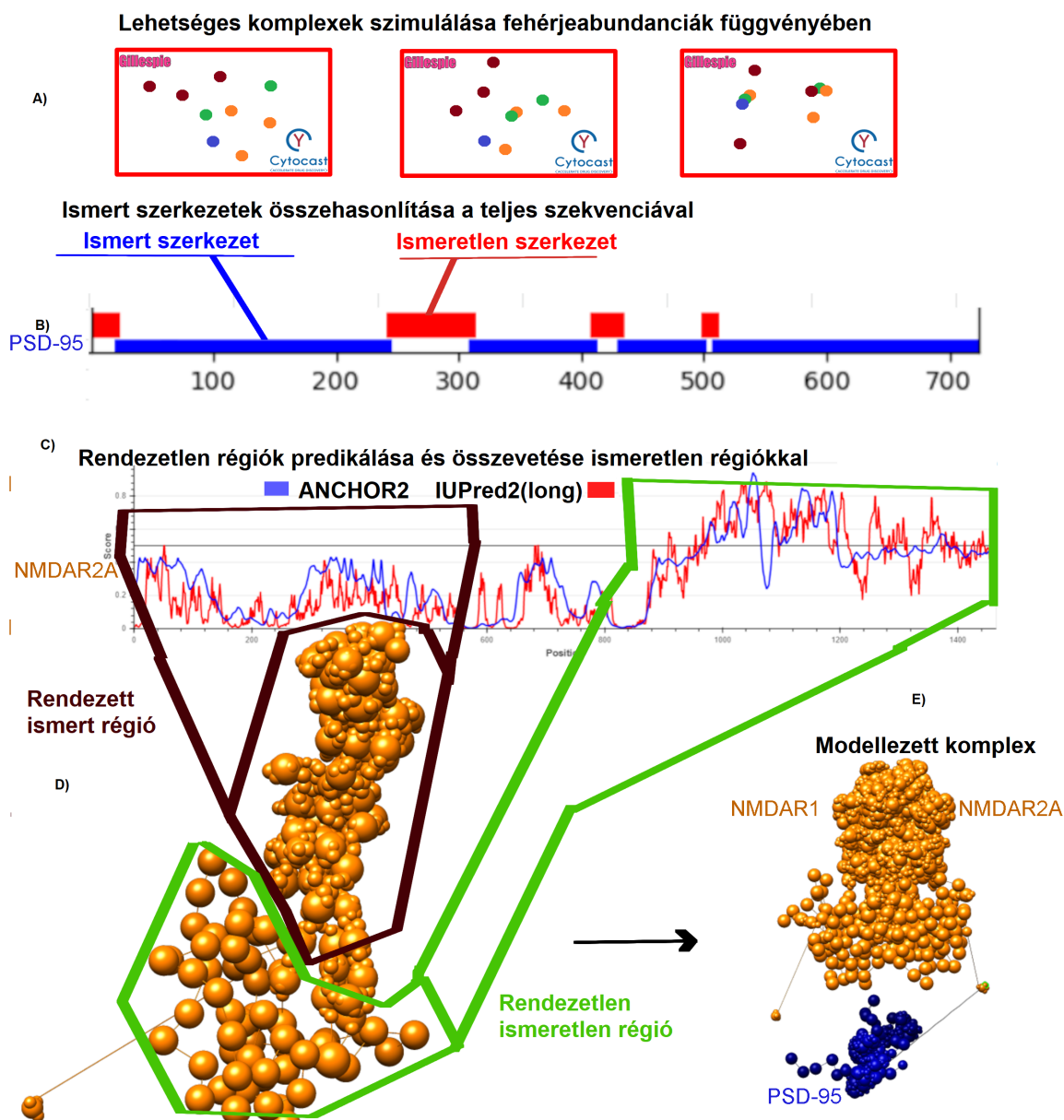
8.16. ábra. A) 2xkx térszerkezet feszültségtérképe. B) egy példa magas feszültségre két aminosav között. Az ábra a szerző saját munkája.

Egy NMDA receptort is modelleztem az elérhető kísérleti szerkezet alapján (lásd 8.17): 6MM9. Ez a struktúra két NMDAR1 (gén: Grin1) és két NMDAR2A (Grin2A) alegységből áll. Az NMDAR2A alegység C-terminálisa képes kötődni a PSD-95 PDZ1-2 tandem doménjéhez.

Az NMDAR2A alegység C-terminális szegmensét (a 800-1400. aminosavakat) kis felbontással, nagyobb mobilitást megengedve (floppy representation) modelleztem, mivel erre a szakaszra rendezetlenség miatt nincs elérhető kísérletes szerkezeti modell. A túl hosszú rendezetlen régió kezelése optimalizálást igényel a komplex modellezés során, és erre még nem találtam teljes mértékben kielégítő megoldást. A komplex szimulációja és a terminális szakasz mozgása a terminális régió gyöngysormodelljének kompaktabbá válását eredményezte, ugyanakkor alig változtatta meg a terminális végének térbeli helyzetét, ezért szükséges a beállítások további tesztelése.

Az emberi fehérjéket reprezentáló térszerkezetek hiányában sokszor csak más fajokból származó szerkezeti modellekre támaszkodhattam melyek alapján a humán szerkezet templát alapú szerkezeti modellezéssel („homológiamodellezés” [176]) állítható elő. Az ilyenfajta modellezés legnagyobb hátránya viszont az, hogy csak hasonló szekvenciájú fehérjék ismert térszerkezeteire hagyatkozhatunk, amelyek nem feltétlenül reprezentálják megfelelően a modellezett fehérjénk valódi térszerkezetét.

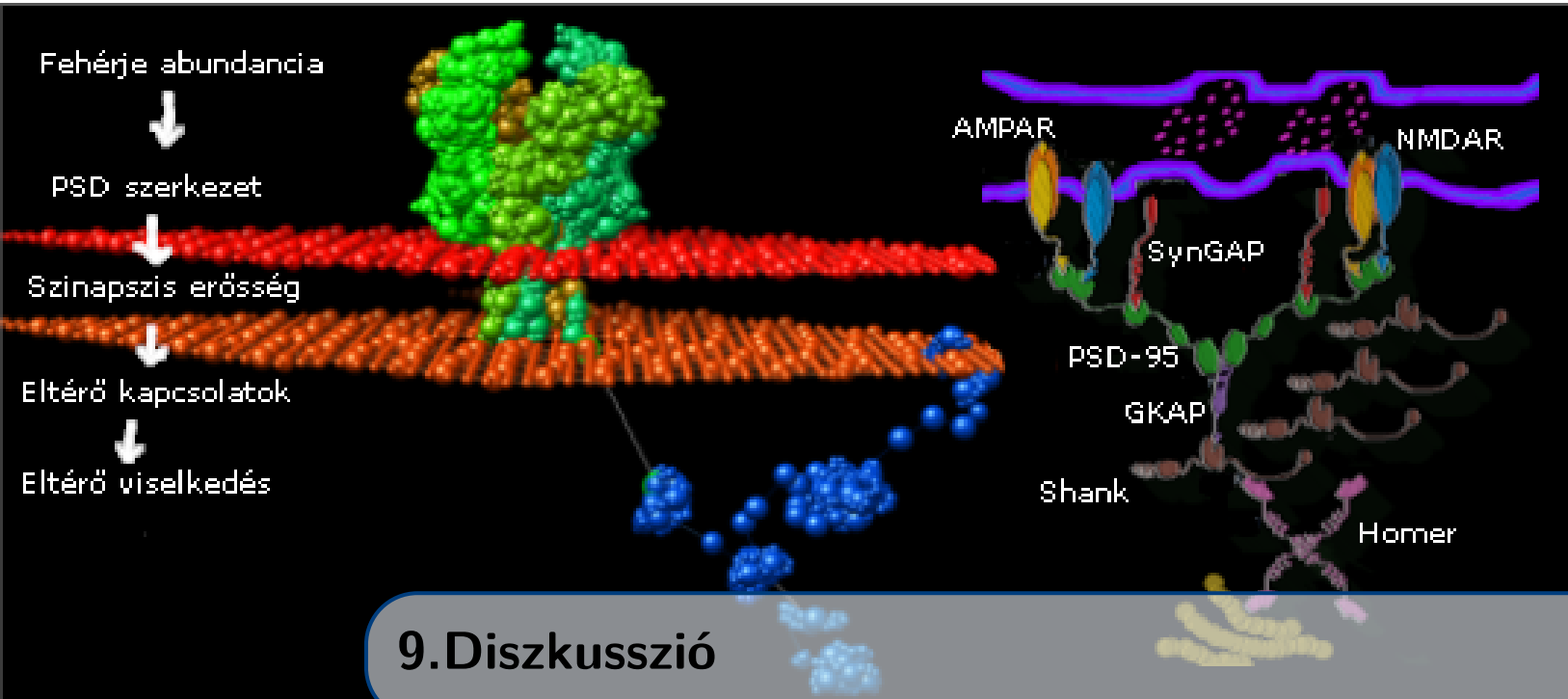
Amennyiben az emberi fehérjére vonatkozó térszerkezet meghatározásra kerül, a komplexek modellezését már könnyebb lesz elvégezni, mert csak ki kell cserélni a modellt az új térszerkezettel. A gyakorlatban az elérhető – tipikusan egérből és patkányból származó fehérjékre vonatkozó – térszerkezeti modellek által lefedett szekvenciák a megfelelő emberi



8.17. ábra. **Az NMDAR/PSD-95 komplex modellezési folyamata** A) A korábbi szimulációk fontos komplexeket mutattak, főleg primer bináris komplexeket. B) A Protein Data Bankban elemeztem az egyes fehérjék ismert szerkezetét (beleértve más fajokból származó nagyon hasonló fehérjéket is). C) Előrejelzéseket készítettem a másodlagos struktúrákra vonatkozóan többek között rendezetlenség szempontjából, hogy megvizsgáljam az ismeretlen részeket is és megkezdtem a bináris komplexek modellezését. D) A rendezetlen régiókat így kisebb felbontásban, de nagyobb flexibilitású gyöngsormodellekkel reprezentáltam az IMP programban. E) Majd a monomer szerkezetek gyöngsormodellejt felhasználva az adott komplexet modelleztem. Az ábra a szerző saját munkája.

szekvenciáktól csak néhány pozícióban különböznek, így nem várható, hogy a térszerkezetek között lényeges különbségek legyenek.

Az, hogy hogyan kezeljük a térszerkezetek változékonyságát, megbízhatóságát, az azért is kulcsfontosságú mert kutatócsoportunk már előző munkák során belátta, hogy a poszt-szinaptikus denzitás fehérjéi számos különböző interakciót képesek végrehajtani, amely a denzitás reorganizációjában kulcsfontosságú szerepet játszhat, melyre Kiss-Tóth Annamáriaék bevezettek egy új leírásra alkalmas fogalmat: Diversity of Potential Interactions (DPI) [177].



9. Diszkusszió

9.1 Megközelítem indoklása

A mindössze hét PSD fehérjét tartalmazó, specifikus térbeli szerveződés nélküli modellünk határozottan egy nagyon leegyszerűsített modell, amely távol áll a posztszinapszis tényleges biológiai összetettségétől. Ezenkívül az egyszerűség kedvéért minden esetben olyan helyzetet veszünk figyelembe, amely homozigóta forgatókönyvnek felel meg, azaz ahol egy adott fehérjének vagy csak a vad típusú, vagy csak a mutáns változata van jelen, de mindkettő egyszerre nincs. Végül, de nem utolsósorban, a kiválasztott mutációnak csak egyetlen jól definiált hatását modelleztük, figyelmen kívül hagyva az olyan lehetséges pleiotróp hatásokat, mint például a többi fehérje expressziós szintjének megváltozása, amint azt pl. számos Shank-mutációnál megfigyelték [178]. Így nem várható, hogy a kapott fehérjekomplex-eloszlásokat közvetlenül össze lehessen hasonlítani az *in vivo* helyzetekkel. Mindezen szempontok elfogadható pontosságú modellezése sokkal több adatot igényelne annál, mint amennyi jelenleg rendelkezésre áll. Mindazonáltal azt állítom, hogy a fő PSD fehérjék és kölcsönhatások jól meghatározott halmazára összpontosító modellrendszerünk elég összetett ahhoz, hogy megragadja a kidolgozott fehérjehálózatok viselkedésének általános aspektusait, miközben kezelhető marad az adatelemzés szempontjából, mivel a modellben a lehetséges fehérjekomplexek száma még nem rendkívül magas. Megfontolásaim szerint a genotípusok és a fenotípusok – itt fenotípus alatt a szinapszisok identitásának meghatározásához funkcionálisan közelebb álló fehérjekomplex-eloszlásokat értve – mechanisztikus összekapcsolása, ahol a különböző genotípusok hasonló fenotípusokhoz vezetnek, csak kísérleti adatok és modellezési megközelítések kombinációja révén lehetséges.

9.1.1 A fehérjegyakorisági adatsorok eloszlásának kritikai áttekintése

A posztszinaptikus denzitás fehérjékének kutatására létrehozandó a modellhez szükséges adatgyűjtés kapcsán több jelentős anomáliára bukkantam, amely megnehezíti az egységes adatbázis létrehozását.

Az első probléma, hogy nem egyértelmű, hogy a vizsgálati alanyok kiválasztása milyen

módon történik. Mint minden statisztikai jellegű vizsgálatra, erre is az optimális kiválasztás a véletlenszerű módszer lenne. Nem lehet tudni, hogy a kutatók ismerik-e az alanyok előzetes kórképét, mely a későbbiekben még igen jelentős szerepet fog játszani vagy játszhat a kapott eredmények értékelésében.

Másrészt gondot jelent, hogy az egyes kutatócsoportok látszólag teljesen önkényesen eltérnek az agy egységes anatómiában használt felosztásától. Egyes agyi részeket együtt kezelnek, míg másokat különválasztanak. Így a kapott mérési eredmények nem, vagy csak részlegesen vethetők össze egymással. Az adatbázis szempontjából ez azt jelenti, hogy az agy anatómiai felosztását reprezentáló kategóriákba különböző esetszámok kerülnek. Míg például az összevont agyi területek mérési eredményeit nem lehet felhasználni, esetleg mindkét területre összevonva be lehet regisztrálni, azonban ez ronthatja az összevetésben a hiba százalékot.

Például a felhasznált adataim között a legtöbb alany esetén az elsődleges motoros kéreg és az elsődleges szenzoros kéreg külön szerepelt, ellenben néhány kivétel esetén a két területet egyben vizsgálták és az adat a két területre együttesen vonatkozik. Ez az összehasonlítást megnehezíti.

Ugyanígy ronthatja a vizsgálati eredményeket az is, hogy bizonyos régiókban találtam kiugró mérési eredményeket. Ez az eredmény nem lehet releváns, ha nincs hozzá további magyarázat. Ugyanis, ha ez csupán egy mérési hiba (hiszen az alany minden agyi területén jelentősen magas PSD-95 fehérje értékeket regisztráltak egységesen), akkor nem használható fel további kutatási célra. Azonban, ha valamilyen valós biológiai, genetikai eltérés okozza ezt (melyről az alanyak nem feltétlenül van tudomása), akkor nagyon is releváns az eredménye, amit nem lehet elhanyagolni.

Gyakorlatilag ez is egy információmegosztási protokollnak tekinthető. Amelyben a következő pontoknak mindenképp szerepelnie kell:

Véletlenszerűen választott alanyok. Előzetes kórkép ismeretének tilalma (befolyásolhatja a kutatót a mérésben - előfeltevése lehet a várható eredményről) - utóelemzésnél vissza lehet tekinteni rá. Az agyi területek anatómiai felosztásának kötelező használata (önkényes eltérések kiküszöbölése). A kiugró (negatív vagy pozitív irányba is) mérési adatok mellett minden esetben megjeleníteni a megismételt mérés eredményét, vagy azt, hogy ez már többszöri mérési eredmény. Illetve épp a fent említett kiugró esetekben az alanyok további vizsgálata szükséges már a kórkép ismeretében, mely alapján esetleg új kutatási eredmények születhetnek az egyes agyi területek, vagy fehérjék funkciójával kapcsolatban, illetve egyes betegségek kapcsán. A legfontosabb viszont a megfelelő dokumentáció és minden egyes mérési adathalmazhoz egy megfelelő cikk elkészítése, amelyben a fenti pontokra is részletesen kitérnek, mert csakis akkor tudjuk kellőképpen elfogadni azokat, mint további vizsgálható és felhasználható adathalmazok. Célszerű egyazon kísérletsorozatból választani minden fehérjére vonatkozó információt - máskülönben a dimenziók és a körülmények eltorzíthatják az eredményeket.

9.2 A PSD-komplexek modelljei – milyen messze vagyunk a valóságtól?

Az mRNS-expressziós adatok lineáris skálázásával becsült fehérjekópiaszámok nem veszik figyelembe a translációs és poszttranszlációs szabályozó hatásokat [179]. Az ilyen hatások a komplexek nagyobb eltérését eredményezhetik az agyi régiók között. Ideális esetben közvetlen fehérjegyakorisági adatokra lenne szükség a valósághűbb szimulációs eredmények

eléréséhez. Ilyen adatok adott képkalkotó – mikroszkópos – technikák segítségével nyerhetőek. Ezek a képkalkotó technikák az mRNS adatokhoz hasonló arányokat mutatnak, mint a validálás, de a dimenziók nem pontos fehérjeszámok, hanem voxelek és intenzitások [180], és az ilyen jellegű adatokból lényegesen kevesebb áll rendelkezésre, mint a nagyobb skálán generálható mRNS expressziós adatokból. A PSD szerveződésének részletes leírása továbbra is kihívást jelent. Bonyolítja a mérete, az alkotó fehérjék száma és variációi, és legfőképpen a változó sztöchiometria és dinamikus jellege. Bár nagy felbontású kísérleti adatok állnak rendelkezésre a bináris komplexekről, ezek jellemzően csak a kölcsönható doméneket és szegmenseket tartalmazzák. Alapvető feltételezésünk, hogy a szimulációk kiegészíthetik a kísérleti adatokat, és jelentős mértékben hozzájárulhatnak a PSD természetének megértéséhez. Az itt bemutatott modellezési megközelítés egy első közelítés, amely elsősorban a fehérjegyakoriságok változékonyságára összpontosít, és csak egy hét PSD fehérjéből álló rendszert vizsgál, és mindegyiknek csak egyetlen izoformája (vagy mutáns változata) található meg benne, így ez a teljes PSD-nek egy erősen leegyszerűsített részhalmaza. Így összetettsége távol áll a PSD tényleges szerveződésétől. Következésképpen eredményeim nem fordíthatók át közvetlenül a PSD-ben ténylegesen megjelenő komplexek valós eloszlására a különböző neuronokban. A pontosabb szimulációkhoz kvantitatív adatokra lenne szükség a közvetlenül fehérjeszinten található mennyiségekre, megfelelő kötési állandókra és a komplexek 3D-s szerveződésének figyelembevételére, valamint a poszt-szinaptikus régió belüli lokalizációjára. Emellett a komponensek dinamikus cseréje, az új fehérjék hozzáadásának térbeli iránya, valamint a fázisszeparáció jelensége mind olyan kérdések, amelyek várhatóan hozzájárulnak a komplexek tényleges eloszlásához *in vivo*.

Szimulációim ugyanakkor már támpontokat adnak a főbb PSD-fehérjék közötti legrelevánsabb összefüggésekre, és rávilágítanak az adott fehérjék hozzáférhetőségében bekövetkezett változásokra. Így bár a modellek nem közelítik meg a valós eset bonyolultságát, megközelítem alkalmas arra, hogy átfogó képet adjunk a PSD főbb szervezőelveiről.

9.3 Összefoglalás és tézispontok

Tézis 1 Irodalmi adatok alapján előkészítettem a poszt-szinaptikus denzitás (PSD) fehérjehálózatának leegyszerűsített modelljét, mely 7 fő fehérjét tartalmaz és bemenetként használható fehérjekomplexek eloszlását szimuláló rendszerbiológiai számításokhoz. Expressziós adatok alapján a hét fehérjét összesen 524 különböző gyakorisági összetételben vizsgáltam a Cytocast szoftver segítségével. [171]

Irodalomkutatás során kísérleti adatokat gyűjtöttem fehérjegyakoriságokra adott alanyok adott agyi régióira lebontva [30]. Mindazonáltal a fehérje gyakoriság mellett annak a valószínűsége, hogy két fehérje találkozik egymással függ a szimulációs rács méretétől és a szimulációs időtől is, nem csak a fehérjék gyakoriságától. Ezért lefuttattam számos szimulációt, amelyek során több különböző metaparamétert is megvizsgáltam és a lehetséges kötések több variációját is figyelembe vettem. Eképpen teszteltem, hogy a szimulációs idő jelentősen nem befolyásolta az eredményeim az elsődleges komplexek arányát illetően. Az azonos metaparaméterekkel és beállításokkal többször is lefuttattam ugyanazt a szimulációt a konfidencia intervallumokat is figyelembe véve. A legtöbb komplex esetén azt kaptam, hogy 40 ismétlés már elegendőnek bizonyult ahhoz, hogy a sztochasztikus ingadozásokat kisimítsam használható mértékben és a konfidenciaintervallum sugarát 1 db komplexre

szűkítsem. Ez azt jelenti, hogy 40 szimuláció kiértékelése során már elég robusztus az eredmény halmaz és elfogadható.

A szimulációk elvégzéséhez számos Python kódot írtam, mind a kísérletes adatokból való bemeneti fájlok legenerálása, mindpedig a kimeneti fájlok kiértékelésére.

Tézis 2 Eljárásokat dolgoztam ki a PSD modell szimulációiból kapott fehérjekomplex-eloszlások összehasonlító elemzésére. Az eljárások lehetővé teszik az egyes fehérjekomplekxek egyértelmű azonosítását és ennek segítségével a különböző szimulációs eredmények összevetését eltérő kiindulási fehérjegyakoriságok és a partnerkötési tulajdonságokat befolyásoló mutációk esetében is. [171]

Szimulációim fő megfigyelése az, hogy a kialakuló komplexek eloszlása bonyolult kapcsolatban áll az őket alkotó fehérjék bőségével. A sok lehetséges kölcsönhatás miatt az adott fehérjék rendelkezésre állása egy bizonyos komplex kialakításához minden lehetséges partnertől függ, jelentős kölcsönös függést biztosítva a különböző fehérje-asszociációk gyakorisága között. Míg a partnerfehérje nagyobb gyakorisága egyszerű esetekben nagyobb kötődési valószínűséget jelenthet, más forgatókönyvekben ez a gyakoriság a gyakoribb partner sok különböző komplexbe történő megkötéséhez vezethet, attól függően, hogy további partnerei elérhetőek-e.

Tézis 3 Megállapítottam, hogy a keletkező fehérjekomplekxek eloszlása és az egyedi fehérjék gyakorisága közötti kapcsolat összetett, csak szimulációkkal térképezhető fel. Nagyon hasonló kiindulási fehérjegyakoriságok is eredményezhetnek jelentősen eltérő fehérjekomplex-eloszlást, melynek a szinaptikus elmélet alapján funkcionális jelentősége lehet. [171], [181]

Az elvégzett szimulációk azt mutatták meg, hogy a felbukkant komplexek bősége nem csak a kezdeti fehérje mennyiségtől függ, hanem a lehetséges kötéseknek is jelentős hatása van rá. Egy bizonyos fehérjét tartalmazó komplex gyakorisága inkább függ a fehérje szomszédainak gyakoriságától, mint magától az adott fehérje gyakoriságától.

Szintén megfigyelhető volt, hogy ha adott egy közepes mennyiségű fehérje, de van egy sokkal gyakoribb kötőpartnere („szomszédja”), akkor, a közepesen gyakori fehérjét tartalmazó komplex megjelenésének valószínűsége magasabb, mintha annak a fehérjének közepesen gyakori szomszédja lenne. Kisebb komplexekből nagyobb komplexek összeállításának mérlegelésekor a közös szomszéd fehérjebősége is fontos, de fordított arányban. Ha a közös szomszéd nagyobb mennyiségben van jelen, akkor kevésbé valószínű, hogy a két fehérje ugyanazon közös szomszédhoz köt (egy másikat hamarabb megtalálnak). A Cytocast kimutatta, hogy ebben az esetben kevesebb az esélye annak, hogy a két komplex, vagy fehérje ugyanabban a komplexben lesz megtalálható.

A szimulációk segítségével azt is bemutattam, hogy a több száz fehérjét tartalmazó szuperkomplekx kialakulásához nem elegendő a Homer1 tetramerizációja, hanem szükséges mellé a Shank3 önasszociációs tulajdonsága is. Mindazonáltal a szuperkomplekxek méretét már befolyásolja a szimulációs idő.

A bemutatott példák a hasonló fehérjekészlettel rendelkező, de a komplexeket eltérő eloszlásban tartalmazó régiókra azt sugallják, hogy a kiválasztott fehérjék lokális szintézise és lebomlása a fehérjekomplekxek olyan mértékű újraeloszlásához vezethet, amely jelentősen megváltoztathatja a szinapszis „azonosságát”. A helyi mRNS transzláció jelenléte a

dendrittüskékben, amelyek többek között PSD-fehérjéket termelnek, jól ismertek, és számos neuronális folyamattal, például késői fázisú LTP-vel [182] összefüggésbe hozhatók. Hasonló komplexeloszlások érhetők el az alkotó fehérjék különböző kombinációival, ami lehetőséget biztosít a funkcionálisan hasonló állapotok elérésére.

A mutációk mechanikus összekapcsolása az általuk okozott fenotípusokkal sokszor nem triviális feladat, különösen akkor, ha a mutáció egy rendkívül összetett fehérjehálózatot zavar meg. Ez a jelenség jól ismert az idegrendszeri fejlődési rendellenességek esetében, ahol a hasonló megfigyelt fenotípusokat számos különböző mutáció okozhatja. Például a Phelan-McDermid-szindrómára vonatkozó közelmúltbeli ajánlás a mögöttes genetikai okra helyezi a hangsúlyt, mivel a fenotípusok nagyrészt nem specifikusak, és általában számos idegrendszeri fejlődési betegségben előfordulhatnak [105].

A mutációk specifikus hatása rejtélyes lehet, különösen akkor, ha számos különböző szövetben jelenlévő fehérjéket érintenek. Ez különösen igaz azokra a sejttípusokra, amelyekben még a főbb partnerek és interakciók is azonosak. A neuronok sokfélesége a poszt-szinaptikus fehérjék eltérő mennyisége szempontjából egyedülálló lehetőséget kínál specifikus mutációk hatásának feltárására egy összetett, de mégis leegyszerűsített többkomponensű rendszerben, amely azonos építőelem-készlettel rendelkezik. A szimuláción alapuló megközelítésünk, amely a fehérjekomplexek képződésére fókuszál, amint azt az alkotó fehérjék bősége és kölcsönhatásai határozzák meg, feltárja, hogy egy adott kölcsönhatást gyengítő mutáció hatása nagymértékben függ a rendszerben lévő összes interakciós partner elérhetőségétől. A kölcsönhatások összetett kölcsönös függése egy olyan forgatókönyvhöz vezet, amelyben a fehérje komplexek eloszlásának általános változásai általában finomak, és csak néhány komplex képződése van jelentősen befolyásolva, és ez a hatás egy jól meghatározott, specifikus fehérjét tartalmazó sejthalmazra korlátozható. Míg a közkeletű bölcsesség azt sugallja, hogy a mutált fehérjéket tartalmazó fehérjekomplexek a leginkább érintettek, és azokban a sejtekben, ahol ezek bőségesek, eredményeink azt mutatják, hogy a legkevesebb érintett fehérjét tartalmazó sejtek is a sérülékenyek közé tartoznak, és a leginkább érintett fehérjetársulások csak közvetve kapcsolódnak a ténylegesen meggyengült interakcióhoz. Szimulációink arra utalnak, hogy a kisagy érintett agyi régió lehet, összhangban a szakirodalomban található eredményekkel [172]–[174].

Tézis 4 Megmutattam, hogy egy hipomorf, azaz egy adott fehérje-fehérje kölcsönhatást csak kismértékben gyengítő mutáció hatása jelentősen függ az adott sejtekre jellemző fehérjegyakoriságoktól, emiatt csak adott sejtekben vagy sejttípusokban okoz szignifikáns változást. Mindemellett az is jellemző, hogy a legnagyobb változást mutató komplexek nem feltétlenül azok, amelyekben változással érintett interakciós partnerek szerepelnek.[169]

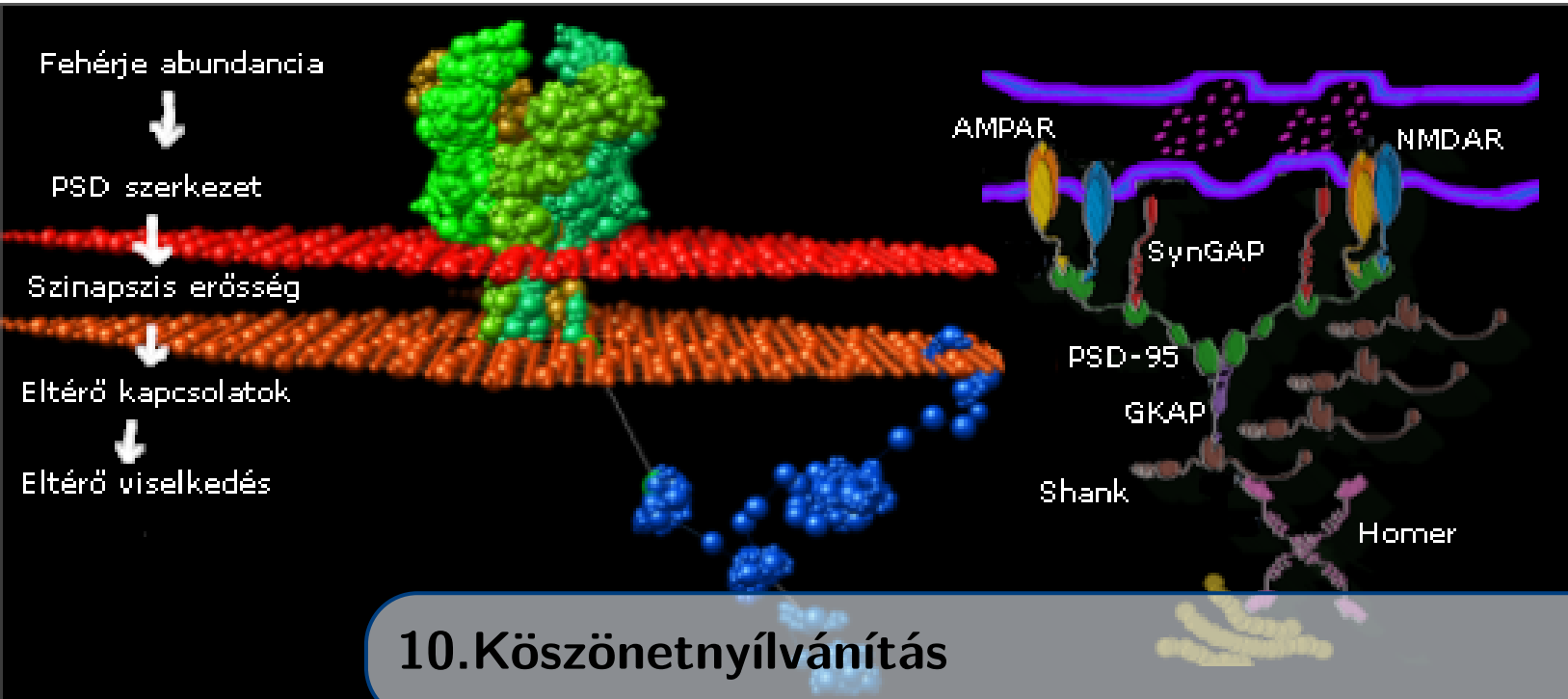
Eredményeim arra is utalnak, hogy a fehérje expressziós szintjét (is) befolyásoló mutációk komolyabb hatást fejtenek ki, mint azok, amelyek mérsékelten zavarják az adott interakciót. Ráadásul a rendszerben a redundancia hatása (pl. Shank1-Shank3) nemcsak azt okozhatja, hogy az egyes izoformák közvetlenül átveszik egymás szerepét, hanem csökkentik a közvetítő hatásokat, pl. itt a legfontosabb komplex gyakoriságának változását mindig a mutánshoz hasonló kötődési mintázatú fehérje jelenléte mérsékelné.

Tézis 5 Kidolgoztam egy munkafolyamatot a posztszinaptikus denzitásra jellemző egyes nagyméretű, többkomponensű fehérjekomplexek lehetséges háromdimenziós szerkezeteinek modellezésére. [175], [183], [184]

Az eljárás két, párhuzamosan futó részből áll, amelyeket később lehet integrálni egy közös modellben - eképpen a Kar két kutatócsoportjának munkáját is összeköti a folyamat. Az egyik irányvonal fehérjekomplexek kialakulásának nagyskálás szimulációja Gillespie algoritmussal, az eljárás másik irányvonala a térszerkezeteken és a szekvenciák annotációin alapszik és figyelembe veszi az egyes fehérjék mint alkotóelemek háromdimenziós jellegét. Ebben az esetben a fehérjékből gyöngysormodellt hozunk létre, ezáltal egy valódi kiterjedéssel rendelkező objektumként tekintünk rájuk. Arra jutottam, hogy a két irányvonal ötvözése nem triviális, de egymásból származó ismeretekkel vegyíthető. Megmutattam, hogy becsülhető a nanoklaszterek térigenye és, hogy a rendezetlen/ismeretlen fehérje régiók kezelése összetett folyamat, mely nagyban befolyásolja a modellt.

9.4 Kutatásom kilátásai: lehetséges hosszútávú irányai és kibontakozása

Az ebben a tézisben foglaltak számos módon kibővíthetőek, és szeretnénk is kibővíteni azokat. Kutatásom fő mozgatórugója, hogy összefüggő, átfogó képet kaphassunk a posztszinaptikus denzitásról, melyhez a kutatócsoport munkája is hozzájárul. Ebbe munkám kezdetleges betekintést nyújt. Eképpen, ahogy halad előre a kutatócsoport, úgy adott rész kutatás eredményeit integrálni lehet a kidolgozott eljárási folyamat segítségével. Pontosíthatóak kötési állandók, új mutációk figyelhetőek meg, szerkezeteken lehet javítani. Megfigyelt sztérikus gátlásokat implementálhatunk. Belső dinamikai információkkal változtathatjuk a domének elérhetőségét - számos konformációt modellezhetünk. Bővíthetjük a fehérjék számát. Kompetíciókat vizsgálhatunk meg. A modell kibővítése kihívásokkal teli folyamatos utazás, amellyel új és eddig kevésbé értett nézőpontból tekinthetünk erre a komplex fehérjehálózatra. Számos új biológiai folyamatot fontolhatunk meg és adhatunk modellünkhöz, hogy közelebb kerüljünk a natív biológiai rendszerekhez - kezdve foszforilációk hatásától, fehérje lebomlástól át külső expressziós szinteket befolyásoló tényezőkhöz. Érdekes kérdés, hogy a cink hogyan hat az oligomerizációra - ténylegesen tud-e létrejönni hosszú Shank-lánc a nanodoméneken keresztül, vagy az ilyen szuperkomplexek - melyek szimulációimban is létrejöhetnek - már nem képezik a posztszinaptikus denzitás aktív részét.



10. Köszönetnyilvánítás

Molekuláris grafika és analízis a the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco által az NIH P41-GM103311 támogatásával fejlesztett UCSF Chimera segítségével történt. A gráfos borítóképek a STRING adatbázis segítségével készültek.

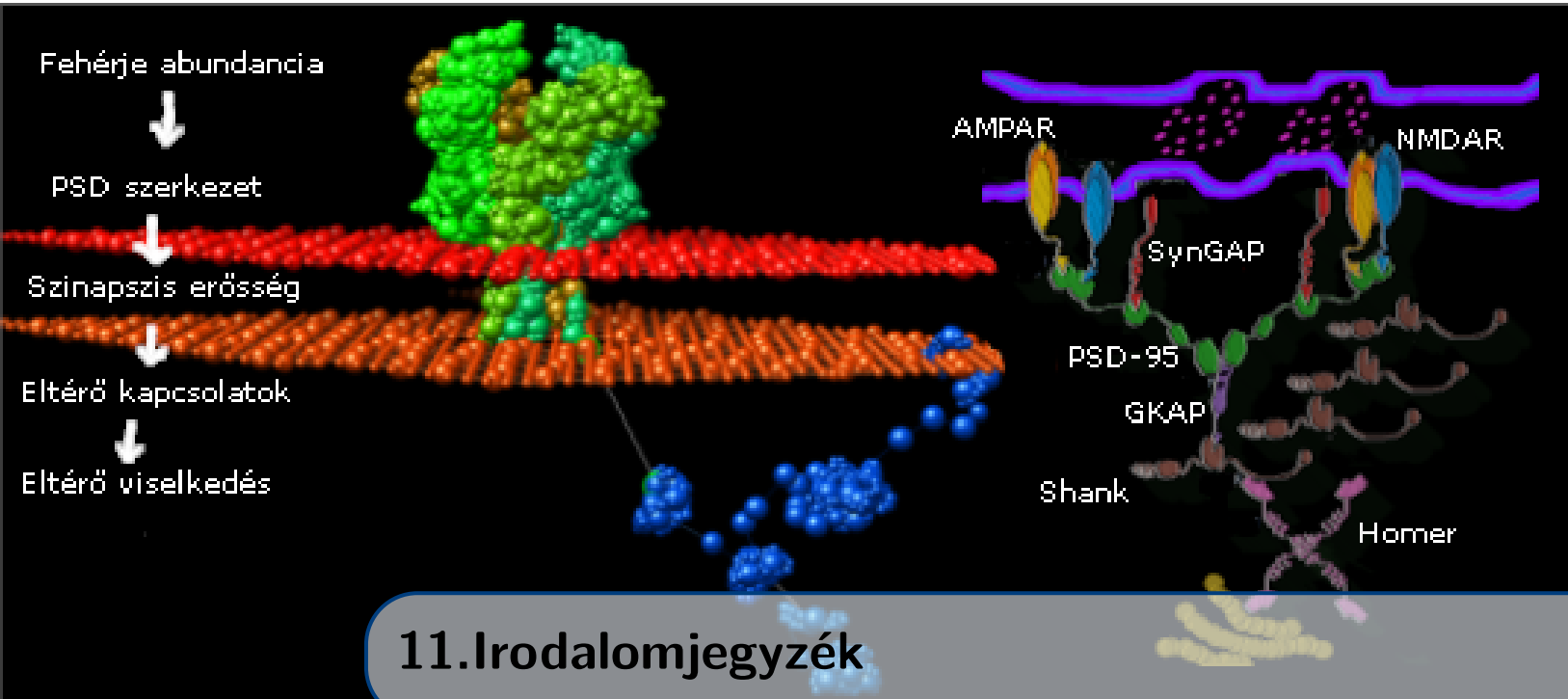
A szerzők elismerik a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal - NKFIH NN124363 számú támogatását. A kutatást az Európai Unió finanszírozta, az Európai Szociális Alap pedig társfinanszírozta EFOP-3.6.2-162017-00013 támogatási szám alatt. A kutatás a Nemzeti Innovációs, Kutatási és Fejlesztési Hivatal OTKA 137947 és TKP2021-EGA-42 számú pályázati támogatásával valósult meg.

Szeretnék köszönetet mondani számos embereknek a kutatásomhoz való hozzájárulásukért. Projektvezetőimnek: Csikász-Nagy Attilának és Gáspári Zoltánnak a felügyeletükért és útmutatásaikért. Köszönöm Komjáti Kornélnak, aki nagy segítségemre volt az IMP alapjainak elsajátításában. Szeretném megköszönni Keömley-Horváth Bence Márk munkáját és, hogy használhattam az általa készített Cytocast szoftvert.

Szeretnék köszönetet mondani Póczik Hunornak a nyelvészeti anglisztikai lektorálásért, amely nagyban hozzájárult a munkám nyelvi minőségének és pontosságának javításához. Hunor szakértelme és alapos munkája nélkülözhetetlen volt a kutatásom során, és hálás vagyok a támogatásáért és a hasznos tanácsaiért.

Továbbá szeretném megköszönni Vida Tivadarné Katinkának a sok sok fáradhatatlan munkát és megértést, amivel segített eligazodni a papírmunka rejtelmében.

Szeretném megköszönni asztaltársaságomnak, azaz a Kerek Lovag Asztalainak a támogatást. Hálás köszönetem kiterjed családomra a sok kitartásukért, türelmükért. Különösen édesanyámnak, Fekete Ágnesnek köszönöm a segítő gondolatait és kérdéseit.



11. Irodalomjegyzék

- [1] R. Melani és N. X. Tritsch, “Inhibitory co-transmission from midbrain dopamine neurons relies on presynaptic GABA uptake”, *Cell Reports*, 39. évfolyam, 3. szám, 110716. oldal, 2022. ápr. DOI: [10.1016/j.celrep.2022.110716](https://doi.org/10.1016/j.celrep.2022.110716). cím: <https://doi.org/10.1016/j.celrep.2022.110716> (hivatkozási oldal 13).
- [2] L. Sancho, M. Contreras és N. J. Allen, “Glia as sculptors of synaptic plasticity”, *Neuroscience Research*, 167. évfolyam, 17–29. oldal, 2021. jún. DOI: [10.1016/j.neures.2020.11.005](https://doi.org/10.1016/j.neures.2020.11.005). cím: <https://doi.org/10.1016/j.neures.2020.11.005> (hivatkozási oldal 14).
- [3] H.-C. Kornau, “Postsynaptic Density/Architecture at Excitatory Synapses 73”, *Reference Module in Neuroscience and Biobehavioral Psychology*, Elsevier, 2017. DOI: [10.1016/b978-0-12-809324-5.02357-9](https://doi.org/10.1016/b978-0-12-809324-5.02357-9). cím: <https://doi.org/10.1016/b978-0-12-809324-5.02357-9> (hivatkozási oldal 14).
- [4] H. Jung, S. Kim, J. Ko és J. W. Um, “Intracellular signaling mechanisms that shape postsynaptic GABAergic synapses”, *Current Opinion in Neurobiology*, 81. évfolyam, 102728. oldal, 2023. aug. DOI: [10.1016/j.conb.2023.102728](https://doi.org/10.1016/j.conb.2023.102728). cím: <https://doi.org/10.1016/j.conb.2023.102728> (hivatkozási oldal 14).
- [5] Y. Yang és J.-J. Liu, “Structural LTP: Signal transduction, actin cytoskeleton reorganization, and membrane remodeling of dendritic spines”, *Current Opinion in Neurobiology*, 74. évfolyam, 102534. oldal, 2022. jún. DOI: [10.1016/j.conb.2022.102534](https://doi.org/10.1016/j.conb.2022.102534). cím: <https://doi.org/10.1016/j.conb.2022.102534> (hivatkozási oldal 14).
- [6] Y. Hayashi, “Molecular mechanism of hippocampal long-term potentiation – Towards multiscale understanding of learning and memory”, *Neuroscience Research*, 175. évfolyam, 3–15. oldal, 2022. febr. DOI: [10.1016/j.neures.2021.08.001](https://doi.org/10.1016/j.neures.2021.08.001). cím: <https://doi.org/10.1016/j.neures.2021.08.001> (hivatkozási oldal 14).

-
- [7] T. M. Boeckers, “The postsynaptic density”, *Cell and Tissue Research*, 326. évfolyam, 2. szám, 409–422. oldal, 2006. júl. DOI: [10.1007/s00441-006-0274-5](https://doi.org/10.1007/s00441-006-0274-5). cím: <https://doi.org/10.1007/s00441-006-0274-5> (hivatkozási oldal 14).
- [8] M. Patterson és R. Yasuda, “Signalling pathways underlying structural plasticity of dendritic spines”, *British Journal of Pharmacology*, 163. évfolyam, 8. szám, 1626–1638. oldal, 2011. júl. DOI: [10.1111/j.1476-5381.2011.01328.x](https://doi.org/10.1111/j.1476-5381.2011.01328.x). cím: <https://doi.org/10.1111/j.1476-5381.2011.01328.x> (hivatkozási oldal 14).
- [9] C. S. Fregozo és M. P. Vega, “Actin-binding proteins and signalling pathways associated with the formation and maintenance of dendritic spines”, *Neurologia (English Edition)*, 27. évfolyam, 7. szám, 421–431. oldal, 2012. szept. DOI: [10.1016/j.nrleng.2011.10.006](https://doi.org/10.1016/j.nrleng.2011.10.006). cím: <https://doi.org/10.1016/j.nrleng.2011.10.006> (hivatkozási oldal 14).
- [10] G. Keilhoff, H. Fansa és G. Wolf, “Neuronal NOS deficiency promotes apoptotic cell death of spinal cord neurons after peripheral nerve transection”, *Nitric Oxide*, 10. évfolyam, 2. szám, 101–111. oldal, 2004. márc. DOI: [10.1016/j.niox.2004.03.008](https://doi.org/10.1016/j.niox.2004.03.008). cím: <https://doi.org/10.1016/j.niox.2004.03.008> (hivatkozási oldal 14).
- [11] R. L. Doser és F. J. Hoerndli, “Regulation of neuronal excitability by reactive oxygen species and calcium signaling: Insights into brain aging”, *Current Research in Neurobiology*, 2. évfolyam, 100012. oldal, 2021. DOI: [10.1016/j.crneur.2021.100012](https://doi.org/10.1016/j.crneur.2021.100012). cím: <https://doi.org/10.1016/j.crneur.2021.100012> (hivatkozási oldal 14).
- [12] A. L. Szabó, A. Sánta, R. Pancsa és Z. Gáspári, “Charged sequence motifs increase the propensity towards liquid–liquid phase separation”, *FEBS Letters*, 596. évfolyam, 8. szám, 1013–1028. oldal, 2022. febr. DOI: [10.1002/1873-3468.14294](https://doi.org/10.1002/1873-3468.14294). cím: <https://doi.org/10.1002/1873-3468.14294> (hivatkozási oldal 14).
- [13] A. Kulkarni és P. Rao, “Synthesis of polymeric nanomaterials for biomedical applications”, *Nanomaterials in Tissue Engineering*, Elsevier, 2013, 27–63. oldal. DOI: [10.1533/9780857097231.1.27](https://doi.org/10.1533/9780857097231.1.27). cím: <https://doi.org/10.1533/9780857097231.1.27> (hivatkozási oldal 14).
- [14] D. Nair, E. Hossy, J. D. Petersen, A. Constals, G. Giannone, D. Choquet és J.-B. Sibarita, “Super-Resolution Imaging Reveals That AMPA Receptors Inside Synapses Are Dynamically Organized in Nanodomains Regulated by PSD95”, *Journal of Neuroscience*, 33. évfolyam, 32. szám, 13204–13224. oldal, 2013. aug. DOI: [10.1523/jneurosci.2381-12.2013](https://doi.org/10.1523/jneurosci.2381-12.2013). cím: <https://doi.org/10.1523/jneurosci.2381-12.2013> (hivatkozási oldal 14).
- [15] W. Droogers és H. MacGillavry, “Plasticity of postsynaptic nanostructure”, *Molecular and Cellular Neuroscience*, 124. évfolyam, 103819. oldal, 2023. márc. DOI: [10.1016/j.mcn.2023.103819](https://doi.org/10.1016/j.mcn.2023.103819). cím: <https://doi.org/10.1016/j.mcn.2023.103819> (hivatkozási oldal 14).
- [16] G. Otto, “Postsynaptic phase separation”, *Nature Reviews Neuroscience*, 22. évfolyam, 7. szám, 386–387. oldal, 2021. jún. DOI: [10.1038/s41583-021-00475-3](https://doi.org/10.1038/s41583-021-00475-3). cím: <https://doi.org/10.1038/s41583-021-00475-3> (hivatkozási oldal 14).

- [17] R. Ortega, S. Roudeau és A. Carmona, “Correlative nano-imaging of metals and proteins in primary neurons by synchrotron X-ray fluorescence and STED super resolution microscopy: Experimental validation”, *Journal of Neuroscience Methods*, 381. évfolyam, 109702. oldal, 2022. nov. DOI: [10.1016/j.jneumeth.2022.109702](https://doi.org/10.1016/j.jneumeth.2022.109702). cím: <https://doi.org/10.1016/j.jneumeth.2022.109702> (hivatkozási oldal 14).
- [18] Z. Feng, X. Chen, M. Zeng és M. Zhang, “Phase separation as a mechanism for assembling dynamic postsynaptic density signalling complexes”, *Current Opinion in Neurobiology*, 57. évfolyam, 1–8. oldal, 2019. aug. DOI: [10.1016/j.conb.2018.12.001](https://doi.org/10.1016/j.conb.2018.12.001). cím: <https://doi.org/10.1016/j.conb.2018.12.001> (hivatkozási oldal 15).
- [19] S. G. Grant, “The Synaptic Theory of Behavior and Brain Disease”, *Cold Spring Harbor Symposia on Quantitative Biology*, 83. évfolyam, 45–56. oldal, 2018. DOI: [10.1101/sqb.2018.83.037887](https://doi.org/10.1101/sqb.2018.83.037887). cím: <https://doi.org/10.1101/sqb.2018.83.037887> (hivatkozási oldal 15).
- [20] S. Franklin és T. M. Vondriska, “Genomes, Proteomes, and the Central Dogma”, *Circulation: Cardiovascular Genetics*, 4. évfolyam, 5. szám, 576–576. oldal, 2011. okt. DOI: [10.1161/circgenetics.110.957795](https://doi.org/10.1161/circgenetics.110.957795). cím: <https://doi.org/10.1161/circgenetics.110.957795> (hivatkozási oldal 15).
- [21] M. Roy, O. Sorokina, N. Skene, C. Simonnet, F. Mazzo, R. Zwart, E. Sher, C. Smith, J. D. Armstrong és S. G. N. Grant, “Proteomic analysis of postsynaptic proteins in regions of the human neocortex”, *Nature Neuroscience*, 21. évfolyam, 1. szám, 130–138. oldal, 2017. dec. DOI: [10.1038/s41593-017-0025-9](https://doi.org/10.1038/s41593-017-0025-9). cím: <https://doi.org/10.1038/s41593-017-0025-9> (hivatkozási oldal 15).
- [22] D. Rosenegger, C. Wright és K. Lukowiak, “A quantitative proteomic analysis of long-term memory”, *Molecular Brain*, 3. évfolyam, 1. szám, 2010. márc. DOI: [10.1186/1756-6606-3-9](https://doi.org/10.1186/1756-6606-3-9). cím: <https://doi.org/10.1186/1756-6606-3-9> (hivatkozási oldal 15).
- [23] J. Griffiths és S. G. Grant, “Synapse pathology in Alzheimer’s disease”, *Seminars in Cell & Developmental Biology*, 139. évfolyam, 13–23. oldal, 2023. ápr. DOI: [10.1016/j.semcdb.2022.05.028](https://doi.org/10.1016/j.semcdb.2022.05.028). cím: <https://doi.org/10.1016/j.semcdb.2022.05.028> (hivatkozási oldal 15).
- [24] A. A. B. Jamjoom, J. Rhodes, P. J. D. Andrews és S. G. N. Grant, “The synapse in traumatic brain injury”, *Brain*, 144. évfolyam, 1. szám, 18–31. oldal, 2020. nov. DOI: [10.1093/brain/awaa321](https://doi.org/10.1093/brain/awaa321). cím: <https://doi.org/10.1093/brain/awaa321> (hivatkozási oldal 15).
- [25] E. Heath-Heckman, S. Yoo, C. Winchell, M. Pellegrino, J. Angstadt, V. B. Lammardo, D. Bautista, F. F. De-Miguel és D. Weisblat, “Transcriptional profiling of identified neurons in leech”, *BMC Genomics*, 22. évfolyam, 1. szám, 2021. márc. DOI: [10.1186/s12864-021-07526-0](https://doi.org/10.1186/s12864-021-07526-0). cím: <https://doi.org/10.1186/s12864-021-07526-0> (hivatkozási oldal 16).
- [26] J. Berg, S. A. Sorensen, J. T. Ting, J. A. Miller, T. Chartrand, A. Buchin, T. E. Bakken, A. Budzillo, N. Dee, S.-L. Ding, N. W. Gouwens, R. D. Hodge, B. Kalmbach, C. Lee, B. R. Lee, L. Alfiler, K. Baker, E. Barkan, A. Beller, K. Berry, D. Bertagnolli, K. Bickley, J. Bomben, T. Braun, K. Brouner, T. Casper, P. Chong, K. Crichton,

- R. Dalley, R. de Frates, T. Desta, S. D. Lee, F. D’Orazi, N. Dotson, T. Egdorf, R. Enstrom, C. Farrell, D. Feng, O. Fong, S. Furdan, A. A. Galakhova, C. Gamlin, A. Gary, A. Glandon, J. Goldy, M. Gorham, N. A. Goriounova, S. Gratiy, L. Graybuck, H. Gu, K. Hadley, N. Hansen, T. S. Heistek, A. M. Henry, D. B. Heyer, D. Hill, C. Hill, M. Hupp, T. Jarsky, S. Kebede, L. Keene, L. Kim, M.-H. Kim, M. Kroll, C. Latimer, B. P. Levi, K. E. Link, M. Mallory, R. Mann, D. Marshall, M. Maxwell, M. McGraw, D. McMillen, E. Melief, E. J. Mertens, L. Mezei, N. Mihut, S. Mok, G. Molnar, A. Mukora, L. Ng, K. Ngo, P. R. Nicovich, J. Nyhus, G. Olah, A. Oldre, V. Omstead, A. Ozsvar, D. Park, H. Peng, T. Pham, C. A. Pom, L. Potekhina, R. Rajanbabu, S. Ransford, D. Reid, C. Rimorin, A. Ruiz, D. Sandman, J. Sulc, S. M. Sunkin, A. Szafer, V. Szemenyei, E. R. Thomsen, M. Tieu, A. Torkelson, J. Trinh, H. Tung, W. Wakeman, F. Waleboer, K. Ward, R. Wilbers, G. Williams, Z. Yao, J.-G. Yoon, C. Anastassiou, A. Arkhipov, P. Barzo, A. Bernard, C. Cobbs, P. C. de Witt Hamer, R. G. Ellenbogen, L. Esposito, M. Ferreira, R. P. Gwinn, M. J. Hawrylycz, P. R. Hof, S. Idema, A. R. Jones, C. D. Keene, A. L. Ko, G. J. Murphy, L. Ng, J. G. Ojemann, A. P. Patel, J. W. Phillips, D. L. Silbergeld, K. Smith, B. Tasic, R. Yuste, I. Segev, C. P. J. de Kock, H. D. Mansvelder, G. Tamas, H. Zeng, C. Koch és E. S. Lein, “Human neocortical expansion involves glutamatergic neuron diversification”, *Nature*, 598. évfolyam, 7879. szám, 151–158. oldal, 2021. okt. DOI: [10.1038/s41586-021-03813-8](https://doi.org/10.1038/s41586-021-03813-8). cím: <https://doi.org/10.1038/s41586-021-03813-8> (hivatkozási oldal 16).
- [27] F. Scala, D. Kobak, M. Bernabucci, Y. Bernaerts, C. R. Cadwell, J. R. Castro, L. Hartmanis, X. Jiang, S. Laturnus, E. Miranda, S. Mulherkar, Z. H. Tan, Z. Yao, H. Zeng, R. Sandberg, P. Berens és A. S. Tolias, “Phenotypic variation of transcriptomic cell types in mouse motor cortex”, *Nature*, 2020. nov. DOI: [10.1038/s41586-020-2907-3](https://doi.org/10.1038/s41586-020-2907-3). cím: <https://doi.org/10.1038/s41586-020-2907-3> (hivatkozási oldal 16).
- [28] M. Zeng, J. Diaz-Alonso, F. Ye, X. Chen, J. Xu, Z. Ji, R. A. Nicoll és M. Zhang, “Phase Separation-Mediated TARP/MAGUK Complex Condensation and AMPA Receptor Synaptic Transmission”, *Neuron*, 104. évfolyam, 3. szám, 529–543.e6., 2019. nov. DOI: [10.1016/j.neuron.2019.08.001](https://doi.org/10.1016/j.neuron.2019.08.001). cím: <https://doi.org/10.1016/j.neuron.2019.08.001> (hivatkozási oldal 16).
- [29] I. Beets, G. Zhang, L. A. Fenk, C. Chen, G. M. Nelson, M.-A. Félix és M. de Bono, “Natural Variation in a Dendritic Scaffold Protein Remodels Experience-Dependent Plasticity by Altering Neuropeptide Expression”, *Neuron*, 105. évfolyam, 1. szám, 106–121.e10., 2020. jan. DOI: [10.1016/j.neuron.2019.10.001](https://doi.org/10.1016/j.neuron.2019.10.001). cím: <https://doi.org/10.1016/j.neuron.2019.10.001> (hivatkozási oldal 16).
- [30] J. A. Miller, S.-L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, A. Ebbert, Z. L. Riley, J. J. Royall, K. Aiona, J. M. Arnold, C. Bennet, D. Bertagnolli, K. Brouner, S. Butler, S. Caldejon, A. Carey, C. Cuhaciyán, R. A. Dalley, N. Dee, T. A. Dolbeare, B. A. C. Facer, D. Feng, T. P. Fliss, G. Gee, J. Goldy, L. Gourley, B. W. Gregor, G. Gu, R. E. Howard, J. M. Jochim, C. L. Kuan, C. Lau, C.-K. Lee, F. Lee, T. A. Lemon, P. Lesnar, B. McMurray, N. Mastan, N. Mosqueda, T. Naluai-Cecchini, N.-K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P. D. Parker, S. E. Parry, A. Stevens, M. Pletikos, M. Reding, K. Roll, D. Sandman, M. Sarreal, S. Shapouri,

- N. V. Shapovalova, E. H. Shen, N. Sjoquist, C. R. Slaughterbeck, M. Smith, A. J. Sodt, D. Williams, L. Zöllei, B. Fischl, M. B. Gerstein, D. H. Geschwind, I. A. Glass, M. J. Hawrylycz, R. F. Hevner, H. Huang, A. R. Jones, J. A. Knowles, P. Levitt, J. W. Phillips, N. Šestan, P. Wohnoutka, C. Dang, A. Bernard, J. G. Hohmann és E. S. Lein, “Transcriptional landscape of the prenatal human brain”, *Nature*, 508. évfolyam, 7495. szám, 199–206. oldal, 2014. ápr. DOI: [10.1038/nature13185](https://doi.org/10.1038/nature13185). cím: <https://doi.org/10.1038/nature13185> (hivatkozási oldalak 16, 56, 61, 81, 91).
- [31] R. Fan és A. Hilfinger, “The effect of microRNA on protein variability and gene expression fidelity”, *Biophysical Journal*, 122. évfolyam, 5. szám, 905–923. oldal, 2023. márc. DOI: [10.1016/j.bpj.2023.01.027](https://doi.org/10.1016/j.bpj.2023.01.027). cím: <https://doi.org/10.1016/j.bpj.2023.01.027> (hivatkozási oldal 16).
- [32] B. Suter, “RNA localization and transport”, *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1861. évfolyam, 10. szám, 938–951. oldal, 2018. okt. DOI: [10.1016/j.bbagr.2018.08.004](https://doi.org/10.1016/j.bbagr.2018.08.004). cím: <https://doi.org/10.1016/j.bbagr.2018.08.004> (hivatkozási oldal 16).
- [33] A. R. Tejedor, A. Garaizar, J. Ramirez és J. R. Espinosa, “RNA-induced modulation of the stability and transport properties of protein biomolecular condensates: on the two sides of the same coin”, *Biophysical Journal*, 121. évfolyam, 3. szám, 356a., 2022. febr. DOI: [10.1016/j.bpj.2021.11.967](https://doi.org/10.1016/j.bpj.2021.11.967). cím: <https://doi.org/10.1016/j.bpj.2021.11.967> (hivatkozási oldal 16).
- [34] A. N. Gasparski, K. Moissoglu, S. Pallikkuth, S. Meydan, N. R. Guydosh és S. Mili, “mRNA location and translation rate determine protein targeting to dual destinations”, *Molecular Cell*, 83. évfolyam, 15. szám, 2726–2738.e9., 2023. aug. DOI: [10.1016/j.molcel.2023.06.036](https://doi.org/10.1016/j.molcel.2023.06.036). cím: <https://doi.org/10.1016/j.molcel.2023.06.036> (hivatkozási oldal 16).
- [35] C. Bédard, A. F. Cisneros, D. Jordan és C. R. Landry, “Correlation between protein abundance and sequence conservation: what do recent experiments say?”, *Current Opinion in Genetics & Development*, 77. évfolyam, 101984. oldal, 2022. dec. DOI: [10.1016/j.gde.2022.101984](https://doi.org/10.1016/j.gde.2022.101984). cím: <https://doi.org/10.1016/j.gde.2022.101984> (hivatkozási oldal 16).
- [36] K. M. Ruff, Y. H. Choi, D. Cox, A. R. Ormsby, Y. Myung, D. B. Ascher, S. E. Radford, R. V. Pappu és D. M. Hatters, “Sequence grammar underlying the unfolding and phase separation of globular proteins”, *Molecular Cell*, 82. évfolyam, 17. szám, 3193–3208.e8., 2022. szept. DOI: [10.1016/j.molcel.2022.06.024](https://doi.org/10.1016/j.molcel.2022.06.024). cím: <https://doi.org/10.1016/j.molcel.2022.06.024> (hivatkozási oldal 16).
- [37] J. A. Tantray, S. Mansoor, R. F. C. Wani és N. U. Nissa, “To identify a protein sample using SDS-PAGE”, *Basic Life Science Methods*, Elsevier, 2023, 87–92. oldal. DOI: [10.1016/b978-0-443-19174-9.00021-0](https://doi.org/10.1016/b978-0-443-19174-9.00021-0). cím: <https://doi.org/10.1016/b978-0-443-19174-9.00021-0> (hivatkozási oldal 16).
- [38] A. J. T. George, “The antibody molecule”, *Diagnostic and Therapeutic Antibodies*, Methods in molecular medicine sorozat, Totowa, NJ: Humana Press, 2000, 1–21. oldal (hivatkozási oldal 16).

-
- [39] J. Webster és D. Oxley, “Protein Identification by MALDI-TOF Mass Spectrometry”, *Chemical Genomics and Proteomics*, Humana Press, 2011. aug., 227–240. oldal. DOI: [10.1007/978-1-61779-349-3_15](https://doi.org/10.1007/978-1-61779-349-3_15). cím: https://doi.org/10.1007/978-1-61779-349-3_15 (hivatkozási oldal 17).
- [40] L. Parker, A. Engel-Hall, K. Drew, G. Steinhardt, D. L. H. Jr, D. Jabon, T. McMurry, D. S. Angulo és S. J. Kron, “Investigating quantitation of phosphorylation using MALDI-TOF mass spectrometry”, *Journal of Mass Spectrometry*, 43. évfolyam, 4. szám, 518–527. oldal, 2007. dec. DOI: [10.1002/jms.1342](https://doi.org/10.1002/jms.1342). cím: <https://doi.org/10.1002/jms.1342> (hivatkozási oldal 17).
- [41] S. W. Hell és M. Kroug, “Ground-state-depletion fluorescence microscopy: A concept for breaking the diffraction resolution limit”, *Applied Physics B Lasers and Optics*, 60. évfolyam, 5. szám, 495–497. oldal, 1995. máj. DOI: [10.1007/bf01081333](https://doi.org/10.1007/bf01081333). cím: <https://doi.org/10.1007/bf01081333> (hivatkozási oldal 17).
- [42] S. Jeong, J. Widengren és J.-C. Lee, “Fluorescent Probes for STED Optical Nanoscopy”, *Nanomaterials*, 12. évfolyam, 1. szám, 21. oldal, 2021. dec. DOI: [10.3390/nano12010021](https://doi.org/10.3390/nano12010021). cím: <https://doi.org/10.3390/nano12010021> (hivatkozási oldal 19).
- [43] J. Dodgson, A. Chessel, S. Cox és R. Carazo-Salas, “Super-Resolution Microscopy: SIM, STED and Localization Microscopy”, 2015. okt., 47–60. oldal, ISBN: 978-3-319-22436-7. DOI: [10.1007/978-3-319-22437-4_3](https://doi.org/10.1007/978-3-319-22437-4_3) (hivatkozási oldal 19).
- [44] S. A. Barnes, L. S. Wijetunge, A. D. Jackson, D. Katsanevaki, E. K. Osterweil, N. H. Komiyama, S. G. Grant, M. F. Bear, U. V. Nägerl, P. C. Kind és D. J. Wyllie, “Convergence of Hippocampal Pathophysiology in Syngap1/syngap1^{fl/fl}/supMice”, *The Journal of Neuroscience*, 35. évfolyam, 45. szám, 15073–15081. oldal, 2015. nov. DOI: [10.1523/jneurosci.1087-15.2015](https://doi.org/10.1523/jneurosci.1087-15.2015). cím: <https://doi.org/10.1523/jneurosci.1087-15.2015> (hivatkozási oldal 19).
- [45] M. J. Broadhead, M. H. Horrocks, F. Zhu, L. Muresan, R. Benavides-Piccione, J. DeFelipe, D. Fricker, M. V. Kopyantsa, R. R. Duncan, D. Klenerman, N. H. Komiyama, S. F. Lee és S. G. N. Grant, “PSD95 nanoclusters are postsynaptic building blocks in hippocampus circuits”, *Scientific Reports*, 6. évfolyam, 1. szám, 2016. ápr. DOI: [10.1038/srep24626](https://doi.org/10.1038/srep24626). cím: <https://doi.org/10.1038/srep24626> (hivatkozási oldal 19).
- [46] A. Dani, B. Huang, J. Bergan, C. Dulac és X. Zhuang, “Superresolution Imaging of Chemical Synapses in the Brain”, *Neuron*, 68. évfolyam, 5. szám, 843–856. oldal, 2010. dec. DOI: [10.1016/j.neuron.2010.11.021](https://doi.org/10.1016/j.neuron.2010.11.021). cím: <https://doi.org/10.1016/j.neuron.2010.11.021> (hivatkozási oldal 20).
- [47] E. Drapeau, N. P. Dorr, G. A. Elder és J. D. Buxbaum, “Absence of strong strain effects in behavioral analyses of Shank3-deficient mice”, *Disease Models & Mechanisms*, 2014. jan. DOI: [10.1242/dmm.013821](https://doi.org/10.1242/dmm.013821). cím: <https://doi.org/10.1242/dmm.013821> (hivatkozási oldal 21).

- [48] J. K. Shepherd, S. S. Grewal, A. Fletcher, D. J. Bill és C. T. Dourish, “Behavioural and pharmacological characterisation of the elevated “zero-maze” as an animal model of anxiety”, *Psychopharmacology*, 116. évfolyam, 1. szám, 56–64. oldal, 1994. szept. DOI: [10.1007/bf02244871](https://doi.org/10.1007/bf02244871). cím: <https://doi.org/10.1007/bf02244871> (hivatkozási oldal 21).
- [49] B. M. Sweis, S. P. Bachour, J. A. Brekke, J. C. Gewirtz, H. Sadeghi-Bazargani, M. Hevesi és A. A. Divani, “A modified beam-walking apparatus for assessment of anxiety in a rodent model of blast traumatic brain injury”, *Behavioural Brain Research*, 296. évfolyam, 149–156. oldal, 2016. jan. DOI: [10.1016/j.bbr.2015.09.015](https://doi.org/10.1016/j.bbr.2015.09.015). cím: <https://doi.org/10.1016/j.bbr.2015.09.015> (hivatkozási oldal 21).
- [50] C. A. Seidel, J. Kubiak, S. Felekyan, D. Rohrbeck, J. J. McCann és M. E. Bowen, “Mapping the Conformational Dynamics of the Scaffold Protein PSD-95”, *Biophysical Journal*, 108. évfolyam, 2. szám, 58a–59a., 2015. jan. DOI: [10.1016/j.bpj.2014.11.354](https://doi.org/10.1016/j.bpj.2014.11.354). cím: <https://doi.org/10.1016/j.bpj.2014.11.354> (hivatkozási oldal 21).
- [51] A. A. Kawale és B. M. Burmann, “Advanced NMR spectroscopy methods to study protein structure and dynamics”, *Advanced Spectroscopic Methods to Study Biomolecular Structure and Dynamics*, Elsevier, 2023, 125–152. oldal. DOI: [10.1016/b978-0-323-99127-8.00010-6](https://doi.org/10.1016/b978-0-323-99127-8.00010-6). cím: <https://doi.org/10.1016/b978-0-323-99127-8.00010-6> (hivatkozási oldal 22).
- [52] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli és D. Hassabis, “Highly accurate protein structure prediction with AlphaFold”, *Nature*, 596. évfolyam, 7873. szám, 583–589. oldal, 2021. júl. DOI: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2). cím: <https://doi.org/10.1038/s41586-021-03819-2> (hivatkozási oldal 22).
- [53] R. J. Read, E. N. Baker, C. S. Bond, E. F. Garman és M. J. van Raaij, “iAlphaFold/i and the future of structural biology”, *IUCrJ*, 10. évfolyam, 4. szám, 377–379. oldal, 2023. jún. DOI: [10.1107/s2052252523004943](https://doi.org/10.1107/s2052252523004943). cím: <https://doi.org/10.1107/s2052252523004943> (hivatkozási oldal 22).
- [54] S. Jin, M. D. Miller, M. Chen, N. P. Schafer, X. Lin, X. Chen, G. N. Phillips és P. G. Wolynes, “Molecular-replacement phasing using predicted protein structures from iAWSEM-Suite/i”, *IUCrJ*, 7. évfolyam, 6. szám, 1168–1178. oldal, 2020. okt. DOI: [10.1107/s2052252520013494](https://doi.org/10.1107/s2052252520013494). cím: <https://doi.org/10.1107/s2052252520013494> (hivatkozási oldal 22).
- [55] T. Terwilliger, D. Liebschner, T. Croll, C. Williams, A. McCoy, B. Poon, P. Afonine, R. Oeffner, J. Richardson, R. Read és P. Adams, *AlphaFold predictions are valuable hypotheses, and accelerate but do not replace experimental structure determination*, 2022. DOI: [10.17863/CAM.92214](https://doi.org/10.17863/CAM.92214). cím: <https://www.repository.cam.ac.uk/handle/1810/344791> (hivatkozási oldal 22).

- [56] F. Pinheiro, J. Santos és S. Ventura, “AlphaFold and the amyloid landscape”, *Journal of Molecular Biology*, 433. évfolyam, 20. szám, 167059. oldal, 2021. okt. DOI: [10.1016/j.jmb.2021.167059](https://doi.org/10.1016/j.jmb.2021.167059). cím: <https://doi.org/10.1016/j.jmb.2021.167059> (hivatkozási oldal 22).
- [57] P. S. F. C. Gomes, D. E. B. Gomes és R. C. Bernardi, “Protein structure prediction in the era of AI: Challenges and limitations when applying to in silico force spectroscopy”, *Frontiers in Bioinformatics*, 2. évfolyam, 2022. okt. DOI: [10.3389/fbinf.2022.983306](https://doi.org/10.3389/fbinf.2022.983306). cím: <https://doi.org/10.3389/fbinf.2022.983306> (hivatkozási oldal 22).
- [58] D. Dudola, A. Hinsenkamp és Z. Gáspári, “Ensemble-Based Analysis of the Dynamic Allostery in the PSD-95 PDZ3 Domain in Relation to the General Variability of PDZ Structures”, *International Journal of Molecular Sciences*, 21. évfolyam, 21. szám, 8348. oldal, 2020. nov. DOI: [10.3390/ijms21218348](https://doi.org/10.3390/ijms21218348). cím: <https://doi.org/10.3390/ijms21218348> (hivatkozási oldal 22).
- [59] B. Kovács, N. Zajácz-Epresi és Z. Gáspári, “Ligand-dependent intra- and interdomain motions in the PDZ12 tandem regulate binding interfaces in postsynaptic density protein-95”, *FEBS Letters*, 594. évfolyam, 5. szám, 887–902. oldal, 2019. okt. DOI: [10.1002/1873-3468.13626](https://doi.org/10.1002/1873-3468.13626). cím: <https://doi.org/10.1002/1873-3468.13626> (hivatkozási oldal 22).
- [60] J. H. Kotaleski és K. T. Blackwell, “Modelling the molecular mechanisms of synaptic plasticity using systems biology approaches”, *Nature Reviews Neuroscience*, 11. évfolyam, 4. szám, 239–251. oldal, 2010. ápr. DOI: [10.1038/nrn2807](https://doi.org/10.1038/nrn2807). cím: <https://doi.org/10.1038/nrn2807> (hivatkozási oldal 22).
- [61] I. Z. Reguly, D. Csércsik, J. Juhász, K. Tornai, Z. Bujtár, G. Horváth, B. Keömley-Horváth, T. Kós, G. Cserey, K. Iván, S. Pongor, G. Szederkényi, G. Röst és A. Csikász-Nagy, “Microsimulation based quantitative analysis of COVID-19 management strategies”, *PLOS Computational Biology*, 18. évfolyam, 1. szám, C. J. Struchiner, szerkesztő, e1009693., 2022. jan. DOI: [10.1371/journal.pcbi.1009693](https://doi.org/10.1371/journal.pcbi.1009693). cím: <https://doi.org/10.1371/journal.pcbi.1009693> (hivatkozási oldal 22).
- [62] B. Webb, S. Viswanath, M. Bonomi, R. Pellarin, C. H. Greenberg, D. Saltzberg és A. Sali, “Integrative structure modeling with the Integrative Modeling Platform”, *Protein Science*, 27. évfolyam, 1. szám, 245–258. oldal, 2017. okt. DOI: [10.1002/pro.3311](https://doi.org/10.1002/pro.3311). cím: <https://doi.org/10.1002/pro.3311> (hivatkozási oldal 23).
- [63] D. Schneidman-Duhovny, M. Hammel és A. Sali, “FoXS: a web server for rapid computation and fitting of SAXS profiles”, *Nucleic Acids Research*, 38. évfolyam, Web Server. szám, W540–W544., 2010. máj. DOI: [10.1093/nar/gkq461](https://doi.org/10.1093/nar/gkq461). cím: <https://doi.org/10.1093/nar/gkq461> (hivatkozási oldal 23).
- [64] D. Schneidman-Duhovny, M. Hammel, J. A. Tainer és A. Sali, “FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles”, *Nucleic Acids Research*, 44. évfolyam, W1. szám, W424–W429., 2016. máj. DOI: [10.1093/nar/gkw389](https://doi.org/10.1093/nar/gkw389). cím: <https://doi.org/10.1093/nar/gkw389> (hivatkozási oldal 23).

- [65] E. Valentini, A. G. Kikhney, G. Previtali, C. M. Jeffries és D. I. Svergun, “SASBDB, a repository for biological small-angle scattering data”, *Nucleic Acids Research*, 43. évfolyam, D1. szám, D357–D363. oldal, 2014. okt. DOI: [10.1093/nar/gku1047](https://doi.org/10.1093/nar/gku1047). cím: <https://doi.org/10.1093/nar/gku1047> (hivatkozási oldal 24).
- [66] A. E. Horner, R. H. Norris, R. McLaren-Jones, L. Alexander, N. H. Komiyama, S. G. N. Grant, J. Nithianantharajah és M. V. Kopanitsa, “Learning and reaction times in mouse touchscreen tests are differentially impacted by mutations in genes encoding postsynaptic interacting proteins scpSYNGAP1/scp, scpNLGN3/scp, scpDLGAP1/scp, scpDLGAP2/scp and scpSHANK2/scp”, *Genes, Brain and Behavior*, 20. évfolyam, 1. szám, 2020. dec. DOI: [10.1111/gbb.12723](https://doi.org/10.1111/gbb.12723). cím: <https://doi.org/10.1111/gbb.12723> (hivatkozási oldal 25).
- [67] A. Sánta, A. Czajlik, G. Batta, B. Péterfia és Z. Gáspári, “Resonance assignment of the Shank1 PDZ domain”, *Biomolecular NMR Assignments*, 16. évfolyam, 1. szám, 121–127. oldal, 2022. jan. DOI: [10.1007/s12104-022-10069-4](https://doi.org/10.1007/s12104-022-10069-4). cím: <https://doi.org/10.1007/s12104-022-10069-4> (hivatkozási oldal 25).
- [68] J. Tamargo, “Pharmacology of cardiac potassium channels”, *Cardiovascular Research*, 62. évfolyam, 1. szám, 9–33. oldal, 2004. ápr. DOI: [10.1016/j.cardiores.2003.12.026](https://doi.org/10.1016/j.cardiores.2003.12.026). cím: <https://doi.org/10.1016/j.cardiores.2003.12.026> (hivatkozási oldal 25).
- [69] Z. D. Bálint Mészáros Gábor Erdős, “IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding”, *Nucleic Acids Research*, 46. évfolyam, W1. szám, W329–W337., 2018. DOI: [doi:10.1093/nar/gky384](https://doi.org/10.1093/nar/gky384). (hivatkozási oldal 25).
- [70] R. B. Nehring, E. Wischmeyer, F. Döring, R. W. Veh, M. Sheng és A. Karschin, “Neuronal Inwardly Rectifying K⁺ Channels Differentially Couple to PDZ Proteins of the PSD-95/SAP90 Family”, *The Journal of Neuroscience*, 20. évfolyam, 1. szám, 156–162. oldal, 2000. jan. DOI: [10.1523/jneurosci.20-01-00156.2000](https://doi.org/10.1523/jneurosci.20-01-00156.2000). cím: <https://doi.org/10.1523/jneurosci.20-01-00156.2000> (hivatkozási oldal 25).
- [71] T. Ishii, K. Moriyoshi, H. Sugihara, K. Sakurada, H. Kadotani, M. Yokoi, C. Akazawa, R. Shigemoto, N. Mizuno és M. Masu, “Molecular characterization of the family of the N-methyl-D-aspartate receptor subunits”, *The Journal of Biological Chemistry*, 268. évfolyam, 4. szám, 2836–2843. oldal, 1993. febr., ISSN: 0021-9258 (hivatkozási oldal 25).
- [72] J. Rudy, *The neurobiology of learning and memory*. New York: Sinauer Associates Oxford University Press, 2021, ISBN: 9781605359342 (hivatkozási oldal 26).
- [73] I. Levitan, *The neuron : cell and molecular biology*. Oxford New York: Oxford University Press, 2015, ISBN: 9780199773893 (hivatkozási oldal 26).
- [74] R. C. Evans, T. Morera-Herreras, Y. Cui, K. Du, T. Sheehan, J. H. Kotaleski, L. Venance és K. T. Blackwell, “The Effects of NMDA Subunit Composition on Calcium Influx and Spike Timing-Dependent Plasticity in Striatal Medium Spiny Neurons”, *PLoS Computational Biology*, 8. évfolyam, 4. szám, A. Morrison, szerkesztő, e1002493., 2012. ápr. DOI: [10.1371/journal.pcbi.1002493](https://doi.org/10.1371/journal.pcbi.1002493). cím: <https://doi.org/10.1371/journal.pcbi.1002493> (hivatkozási oldal 26).

-
- [75] T. Yamakura és K. Shimoji, “Subunit- and site-specific pharmacology of the NMDA receptor channel”, *Progress in Neurobiology*, 59. évfolyam, 3. szám, 279–298. oldal, 1999. okt. DOI: [10.1016/s0301-0082\(99\)00007-6](https://doi.org/10.1016/s0301-0082(99)00007-6). cím: [https://doi.org/10.1016/s0301-0082\(99\)00007-6](https://doi.org/10.1016/s0301-0082(99)00007-6) (hivatkozási oldal 26).
- [76] A. Kuryatov, “Mutational analysis of the glycine-binding site of the NMDA receptor: Structural similarity with bacterial amino acid-binding proteins”, *Neuron*, 12. évfolyam, 6. szám, 1291–1300. oldal, 1994. jún. DOI: [10.1016/0896-6273\(94\)90445-6](https://doi.org/10.1016/0896-6273(94)90445-6). cím: [https://doi.org/10.1016/0896-6273\(94\)90445-6](https://doi.org/10.1016/0896-6273(94)90445-6) (hivatkozási oldal 26).
- [77] T. Ishii, K. Moriyoshi, H. Sugihara, K. Sakurada, H. Kadotani, M. Yokoi, C. Akazawa, R. Shigemoto, N. Mizuno és M. Masu, “Molecular characterization of the family of the N-methyl-D-aspartate receptor subunits”, *J Biol Chem*, 268. évfolyam, 4. szám, 2836–2843. oldal, 1993. febr. (hivatkozási oldal 26).
- [78] Z. A. Ataman, L. Gakhar, B. R. Sorensen, J. W. Hell és M. A. Shea, “The NMDA Receptor NR1 C1 Region Bound to Calmodulin: Structural Insights into Functional Differences between Homologous Domains”, *Structure*, 15. évfolyam, 12. szám, 1603–1617. oldal, 2007. dec. DOI: [10.1016/j.str.2007.10.012](https://doi.org/10.1016/j.str.2007.10.012). cím: <https://doi.org/10.1016/j.str.2007.10.012> (hivatkozási oldal 26).
- [79] M. Passafaro, C. Sala, M. Niethammer és M. Sheng, “Microtubule binding by CRIPT and its potential role in the synaptic clustering of PSD-95”, *Nature Neuroscience*, 2. évfolyam, 12. szám, 1063–1069. oldal, 1999. dec. DOI: [10.1038/15990](https://doi.org/10.1038/15990). cím: <https://doi.org/10.1038/15990> (hivatkozási oldalak 26, 28).
- [80] X. Chen, C. D. Nelson, X. Li, C. A. Winters, R. Azzam, A. A. Sousa, R. D. Leapman, H. Gainer, M. Sheng és T. S. Reese, “PSD-95 Is Required to Sustain the Molecular Organization of the Postsynaptic Density”, *Journal of Neuroscience*, 31. évfolyam, 17. szám, 6329–6338. oldal, 2011. ápr. DOI: [10.1523/jneurosci.5968-10.2011](https://doi.org/10.1523/jneurosci.5968-10.2011). cím: <https://doi.org/10.1523/jneurosci.5968-10.2011> (hivatkozási oldalak 26, 27).
- [81] A. Kamalova és T. Nakagawa, “AMPA receptor structure and auxiliary subunits”, *The Journal of Physiology*, 599. évfolyam, 2. szám, 453–469. oldal, 2020. febr. DOI: [10.1113/jp278701](https://doi.org/10.1113/jp278701). cím: <https://doi.org/10.1113/jp278701> (hivatkozási oldal 26).
- [82] M. H. Hastings és H.-Y. Man, “Synaptic Capture of Laterally Diffusing AMPA Receptors – An Idea That Stuck”, *Trends in Neurosciences*, 41. évfolyam, 6. szám, 330–332. oldal, 2018. jún. DOI: [10.1016/j.tins.2018.03.016](https://doi.org/10.1016/j.tins.2018.03.016). cím: <https://doi.org/10.1016/j.tins.2018.03.016> (hivatkozási oldal 26).
- [83] P.-W. Liu, T. Hosokawa és Y. Hayashi, “Regulation of synaptic nanodomain by liquid–liquid phase separation: A novel mechanism of synaptic plasticity”, *Current Opinion in Neurobiology*, 69. évfolyam, 84–92. oldal, 2021. aug. DOI: [10.1016/j.conb.2021.02.004](https://doi.org/10.1016/j.conb.2021.02.004). cím: <https://doi.org/10.1016/j.conb.2021.02.004> (hivatkozási oldal 27).
- [84] P. Opazo, S. Labrecque, C. M. Tigaret, A. Frouin, P. W. Wiseman, P. D. Koninck és D. Choquet, “CaMKII Triggers the Diffusional Trapping of Surface AMPARs through Phosphorylation of Stargazin”, *Neuron*, 67. évfolyam, 2. szám, 239–252. oldal,

2010. júl. DOI: [10.1016/j.neuron.2010.06.007](https://doi.org/10.1016/j.neuron.2010.06.007). cím: <https://doi.org/10.1016/j.neuron.2010.06.007> (hivatkozási oldal 27).
- [85] A. J. Borgdorff és D. Choquet, “Regulation of AMPA receptor lateral movements”, *Nature*, 417. évfolyam, 6889. szám, 649–653. oldal, 2002. jún. DOI: [10.1038/nature00780](https://doi.org/10.1038/nature00780). cím: <https://doi.org/10.1038/nature00780> (hivatkozási oldal 27).
- [86] A. B. Young, “Huntingtin in health and disease”, *Journal of Clinical Investigation*, 111. évfolyam, 3. szám, 299–302. oldal, 2003. febr. DOI: [10.1172/jci17742](https://doi.org/10.1172/jci17742). cím: <https://doi.org/10.1172/jci17742> (hivatkozási oldal 27).
- [87] D. M. Chetkovich, R. C. Bunn, S.-H. Kuo, Y. Kawasaki, M. Kohwi és D. S. Brecht, “Postsynaptic Targeting of Alternative Postsynaptic Density-95 Isoforms by Distinct Mechanisms”, *The Journal of Neuroscience*, 22. évfolyam, 15. szám, 6415–6425. oldal, 2002. aug. DOI: [10.1523/jneurosci.22-15-06415.2002](https://doi.org/10.1523/jneurosci.22-15-06415.2002). cím: <https://doi.org/10.1523/jneurosci.22-15-06415.2002> (hivatkozási oldal 28).
- [88] A. Buszka, A. Pytyś, D. Colvin, J. Włodarczyk és T. Wójtowicz, “S-Palmitoylation of Synaptic Proteins in Neuronal Plasticity in Normal and Pathological Brains”, *Cells*, 12. évfolyam, 3. szám, 387. oldal, 2023. jan. DOI: [10.3390/cells12030387](https://doi.org/10.3390/cells12030387). cím: <https://doi.org/10.3390/cells12030387> (hivatkozási oldal 28).
- [89] Y. Araki, I. Hong, T. R. Gamache, S. Ju, L. Collado-Torres, J. H. Shin és R. L. Huganir, “SynGAP isoforms differentially regulate synaptic plasticity and dendritic development”, *eLife*, 9. évfolyam, 2020. jún. DOI: [10.7554/elife.56273](https://doi.org/10.7554/elife.56273). cím: <https://doi.org/10.7554/elife.56273> (hivatkozási oldal 28).
- [90] H. J. Carlisle, P. Manzerra, E. Marcora és M. B. Kennedy, “SynGAP Regulates Steady-State and Activity-Dependent Phosphorylation of Cofilin”, *The Journal of Neuroscience*, 28. évfolyam, 50. szám, 13673–13683. oldal, 2008. dec. DOI: [10.1523/jneurosci.4695-08.2008](https://doi.org/10.1523/jneurosci.4695-08.2008). cím: <https://doi.org/10.1523/jneurosci.4695-08.2008> (hivatkozási oldal 28).
- [91] X. Ye és T. J. Carew, “Small G Protein Signaling in Neuronal Plasticity and Memory Formation: The Specific Role of Ras Family Proteins”, *Neuron*, 68. évfolyam, 3. szám, 340–361. oldal, 2010. nov. DOI: [10.1016/j.neuron.2010.09.013](https://doi.org/10.1016/j.neuron.2010.09.013). cím: <https://doi.org/10.1016/j.neuron.2010.09.013> (hivatkozási oldal 28).
- [92] J. H. Kim, H.-K. Lee, K. Takamiya és R. L. Huganir, “The Role of Synaptic GTPase-Activating Protein in Neuronal Development and Synaptic Plasticity”, *The Journal of Neuroscience*, 23. évfolyam, 4. szám, 1119–1124. oldal, 2003. febr. DOI: [10.1523/jneurosci.23-04-01119.2003](https://doi.org/10.1523/jneurosci.23-04-01119.2003). cím: <https://doi.org/10.1523/jneurosci.23-04-01119.2003> (hivatkozási oldal 28).
- [93] S. M. O’Mara, S. Commins és M. Anderson, “Synaptic plasticity in the hippocampal area CA1-subiculum projection: Implications for theories of memory”, *Hippocampus*, 10. évfolyam, 4. szám, 447–456. oldal, 2000. DOI: [10.1002/1098-1063\(2000\)10:4<447::aid-hipo11>3.0.co;2-2](https://doi.org/10.1002/1098-1063(2000)10:4<447::aid-hipo11>3.0.co;2-2). cím: [https://doi.org/10.1002/1098-1063\(2000\)10:4%3C447::aid-hipo11%3E3.0.co;2-2](https://doi.org/10.1002/1098-1063(2000)10:4%3C447::aid-hipo11%3E3.0.co;2-2) (hivatkozási oldal 28).

-
- [94] S. M. Shin, N. Zhang, J. Hansen, N. Z. Gerges, D. T. S. Pak, M. Sheng és S. H. Lee, “GKAP orchestrates activity-dependent postsynaptic protein remodeling and homeostatic scaling”, *Nature Neuroscience*, 15. évfolyam, 12. szám, 1655–1666. oldal, 2012. nov. DOI: [10.1038/nn.3259](https://doi.org/10.1038/nn.3259). cím: <https://doi.org/10.1038/nn.3259> (hivatkozási oldalak 28, 29).
- [95] E. Kim, S. Naisbitt, Y.-P. Hsueh, A. Rao, A. Rothschild, A. M. Craig és M. Sheng, “GKAP, a Novel Synaptic Protein That Interacts with the Guanylate Kinase-like Domain of the PSD-95/SAP90 Family of Channel Clustering Molecules”, *Journal of Cell Biology*, 136. évfolyam, 3. szám, 669–678. oldal, 1997. febr. DOI: [10.1083/jcb.136.3.669](https://doi.org/10.1083/jcb.136.3.669). cím: <https://doi.org/10.1083/jcb.136.3.669> (hivatkozási oldal 29).
- [96] J. C. Tu, B. Xiao, S. Naisbitt, J. P. Yuan, R. S. Petralia, P. Brakeman, A. Doan, V. K. Aakalu, A. A. Lanahan, M. Sheng és P. F. Worley, “Coupling of mGluR/Homer and PSD-95 Complexes by the Shank Family of Postsynaptic Density Proteins”, *Neuron*, 23. évfolyam, 3. szám, 583–592. oldal, 1999. júl. DOI: [10.1016/s0896-6273\(00\)80810-7](https://doi.org/10.1016/s0896-6273(00)80810-7). cím: [https://doi.org/10.1016/s0896-6273\(00\)80810-7](https://doi.org/10.1016/s0896-6273(00)80810-7) (hivatkozási oldalak 29, 30).
- [97] S. Naisbitt, E. Kim, J. C. Tu, B. Xiao, C. Sala, J. Valtschanoff, R. J. Weinberg, P. F. Worley és M. Sheng, “Shank, a Novel Family of Postsynaptic Density Proteins that Binds to the NMDA Receptor/PSD-95/GKAP Complex and Cortactin”, *Neuron*, 23. évfolyam, 3. szám, 569–582. oldal, 1999. júl. DOI: [10.1016/s0896-6273\(00\)80809-0](https://doi.org/10.1016/s0896-6273(00)80809-0). cím: [https://doi.org/10.1016/s0896-6273\(00\)80809-0](https://doi.org/10.1016/s0896-6273(00)80809-0) (hivatkozási oldal 29).
- [98] A. Mossa, F. Giona, J. Pagano, C. Sala és C. Verpelli, “SHANK genes in autism: Defining therapeutic targets”, *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 84. évfolyam, 416–423. oldal, 2018. jún. DOI: [10.1016/j.pnpbp.2017.11.019](https://doi.org/10.1016/j.pnpbp.2017.11.019). cím: <https://doi.org/10.1016/j.pnpbp.2017.11.019> (hivatkozási oldal 29).
- [99] P. Monteiro és G. Feng, “SHANK proteins: roles at the synapse and in autism spectrum disorder”, *Nature Reviews Neuroscience*, 18. évfolyam, 3. szám, 147–157. oldal, 2017. febr. DOI: [10.1038/nrn.2016.183](https://doi.org/10.1038/nrn.2016.183). cím: <https://doi.org/10.1038/nrn.2016.183> (hivatkozási oldal 29).
- [100] M. H. Arons, K. Lee, C. J. Thynne, S. A. Kim, C. Schob, S. Kindler, J. M. Montgomery és C. C. Garner, “Shank3 Is Part of a Zinc-Sensitive Signaling System That Regulates Excitatory Synaptic Strength”, *Journal of Neuroscience*, 36. évfolyam, 35. szám, 9124–9134. oldal, 2016. aug. DOI: [10.1523/jneurosci.0116-16.2016](https://doi.org/10.1523/jneurosci.0116-16.2016). cím: <https://doi.org/10.1523/jneurosci.0116-16.2016> (hivatkozási oldal 29).
- [101] M. K. Baron, “An Architectural Framework That May Lie at the Core of the Postsynaptic Density”, *Science*, 311. évfolyam, 5760. szám, 531–535. oldal, 2006. jan. DOI: [10.1126/science.1118995](https://doi.org/10.1126/science.1118995). cím: <https://doi.org/10.1126/science.1118995> (hivatkozási oldalak 29, 65).
- [102] P. Kursula, “Shanks — multidomain molecular scaffolds of the postsynaptic density”, *Current Opinion in Structural Biology*, 54. évfolyam, 122–128. oldal, 2019, Folding and Binding • Proteins:3D structures, ISSN: 0959-440X. DOI: <https://doi.org/10.1016/j.sbi.2019.01.007>. cím: <https://www.sciencedirect.com/science/article/pii/S0959440X18300666> (hivatkozási oldal 29).

- [103] J. P. Delling és T. M. Boeckers, “Comparison of SHANK3 deficiency in animal models: phenotypes, treatment strategies, and translational implications”, *Journal of Neurodevelopmental Disorders*, 13. évfolyam, 1. szám, 2021. nov. DOI: [10.1186/s11689-021-09397-8](https://doi.org/10.1186/s11689-021-09397-8). cím: <https://doi.org/10.1186/s11689-021-09397-8> (hivatkozási oldal 29).
- [104] C. S. Leblond, C. Nava, A. Polge, J. Gauthier, G. Huguet, S. Lumbroso, F. Giuliano, C. Stordeur, C. Depienne, K. Mouzat, D. Pinto, J. Howe, N. Lemièrre, C. M. Durand, J. Guibert, E. Ey, R. Toro, H. Peyre, A. Mathieu, F. Amsellem, M. Rastam, I. C. Gillberg, G. A. Rappold, R. Holt, A. P. Monaco, E. Maestrini, P. Galan, D. Heron, A. Jacqueline, A. Afenjar, A. Rastetter, A. Brice, F. Devillard, B. Assouline, F. Laffargue, J. Lespinasse, J. Chiesa, F. Rivier, D. Bonneau, B. Regnault, D. Zelenika, M. Delepine, M. Lathrop, D. Sanlaville, C. Schluth-Bolard, P. Edery, L. Perrin, A. C. Tabet, M. J. Schmeisser, T. M. Boeckers, M. Coleman, D. Sato, P. Szatmari, S. W. Scherer, G. A. Rouleau, C. Betancur, M. Leboyer, C. Gillberg, R. Delorme és T. Bourgeron, “Meta-analysis of SHANK Mutations in Autism Spectrum Disorders: A Gradient of Severity in Cognitive Impairments”, *PLoS Genetics*, 10. évfolyam, 9. szám, G. S. Barsh, szerkesztő, e1004580., 2014. szept. DOI: [10.1371/journal.pgen.1004580](https://doi.org/10.1371/journal.pgen.1004580). cím: <https://doi.org/10.1371/journal.pgen.1004580> (hivatkozási oldalak 29, 64).
- [105] M. Schön, P. Lapunzina, J. Nevado, T. Mattina, C. Gunnarsson, K. Hadzsiev, C. VerPELLI, T. Bourgeron, S. Jesse, C. M. van Ravenswaaij-Arts és R. C. Hennekam, “Definition and clinical variability of SHANK3-related Phelan-McDermid syndrome”, *European Journal of Medical Genetics*, 66. évfolyam, 7. szám, 104754. oldal, 2023. júl. DOI: [10.1016/j.ejmg.2023.104754](https://doi.org/10.1016/j.ejmg.2023.104754). cím: <https://doi.org/10.1016/j.ejmg.2023.104754> (hivatkozási oldalak 29, 93).
- [106] C. Betancur és J. D. Buxbaum, “SHANK3 haploinsufficiency: a “common” but underdiagnosed highly penetrant monogenic cause of autism spectrum disorders”, *Molecular Autism*, 4. évfolyam, 1. szám, 17. oldal, 2013. DOI: [10.1186/2040-2392-4-17](https://doi.org/10.1186/2040-2392-4-17). cím: <https://doi.org/10.1186/2040-2392-4-17> (hivatkozási oldal 29).
- [107] M. Ali, M. M. McAuley, S. Lüchow, S. Knapp, A. C. Joerger és Y. Ivarsson, “Integrated analysis of Shank1 PDZ interactions with C-terminal and internal binding motifs”, *Current Research in Structural Biology*, 3. évfolyam, 41–50. oldal, 2021. DOI: [10.1016/j.crstbi.2021.01.001](https://doi.org/10.1016/j.crstbi.2021.01.001). cím: <https://doi.org/10.1016/j.crstbi.2021.01.001> (hivatkozási oldal 29).
- [108] S. Yoon, N. H. Piguel, N. Khalatyan, L. E. Dionisio, J. N. Savas és P. Penzes, “Homer1 promotes dendritic spine growth through ankyrin-G and its loss reshapes the synaptic proteome”, *Molecular Psychiatry*, 26. évfolyam, 6. szám, 1775–1789. oldal, 2021. jan. DOI: [10.1038/s41380-020-00991-1](https://doi.org/10.1038/s41380-020-00991-1). cím: <https://doi.org/10.1038/s41380-020-00991-1> (hivatkozási oldal 30).
- [109] X. Chen, B. Jia, S. Zhu és M. Zhang, “Phase separation-mediated actin bundling by the postsynaptic density condensates”, *eLife*, 12. évfolyam, 2023. jún. DOI: [10.7554/eLife.84446](https://doi.org/10.7554/eLife.84446). cím: <https://doi.org/10.7554/eLife.84446> (hivatkozási oldal 30).

-
- [110] A. de Bartolomeis, A. Barone, E. F. Buonaguro, C. Tomasetti, L. Vellucci és F. Iasevoli, “The Homer1 family of proteins at the crossroad of dopamine-glutamate signaling: An emerging molecular “Lego” in the pathophysiology of psychiatric disorders. A systematic review and translational insight”, *Neuroscience & Biobehavioral Reviews*, 136. évfolyam, 104596. oldal, 2022. máj. DOI: [10.1016/j.neubiorev.2022.104596](https://doi.org/10.1016/j.neubiorev.2022.104596). cím: <https://doi.org/10.1016/j.neubiorev.2022.104596> (hivatkozási oldal 30).
- [111] J. Lu, T. D. Helton, T. A. Blanpied, B. Rácz, T. M. Newpher, R. J. Weinberg és M. D. Ehlers, “Postsynaptic Positioning of Endocytic Zones and AMPA Receptor Cycling by Physical Coupling of Dynamin-3 to Homer”, *Neuron*, 55. évfolyam, 6. szám, 874–889. oldal, 2007. szept. DOI: [10.1016/j.neuron.2007.06.041](https://doi.org/10.1016/j.neuron.2007.06.041). cím: <https://doi.org/10.1016/j.neuron.2007.06.041> (hivatkozási oldal 30).
- [112] J.-H. Hu, L. Yang, P. J. Kammermeier, C. G. Moore, P. R. Brakeman, J. Tu, S. Yu, R. S. Petralia, Z. Li, P.-W. Zhang, J. M. Park, X. Dong, B. Xiao és P. F. Worley, “Preso1 dynamically regulates group I metabotropic glutamate receptors”, *Nature Neuroscience*, 15. évfolyam, 6. szám, 836–844. oldal, 2012. máj. DOI: [10.1038/nn.3103](https://doi.org/10.1038/nn.3103). cím: <https://doi.org/10.1038/nn.3103> (hivatkozási oldal 30).
- [113] B. Xiao, J. C. Tu és P. F. Worley, “Homer: a link between neural activity and glutamate receptor function”, *Current Opinion in Neurobiology*, 10. évfolyam, 3. szám, 370–374. oldal, 2000. jún. DOI: [10.1016/s0959-4388\(00\)00087-8](https://doi.org/10.1016/s0959-4388(00)00087-8). cím: [https://doi.org/10.1016/s0959-4388\(00\)00087-8](https://doi.org/10.1016/s0959-4388(00)00087-8) (hivatkozási oldal 30).
- [114] G. H. Diering, R. S. Nirujogi, R. H. Roth, P. F. Worley, A. Pandey és R. L. Huganir, “Homer1a drives homeostatic scaling-down of excitatory synapses during sleep”, *Science*, 355. évfolyam, 6324. szám, 511–515. oldal, 2017. febr. DOI: [10.1126/science.aai8355](https://doi.org/10.1126/science.aai8355). cím: <https://doi.org/10.1126/science.aai8355> (hivatkozási oldal 30).
- [115] V. V. Reshetnikov és N. P. Bondar, “The Role of Stress-Induced Changes of Homer1 Expression in Stress Susceptibility”, *Biochemistry (Moscow)*, 86. évfolyam, 6. szám, 613–626. oldal, 2021. jún. DOI: [10.1134/s0006297921060018](https://doi.org/10.1134/s0006297921060018). cím: <https://doi.org/10.1134/s0006297921060018> (hivatkozási oldal 30).
- [116] “UniProt: a worldwide hub of protein knowledge”, *Nucleic Acids Research*, 47. évfolyam, D1. szám, D506–D515. oldal, 2018. nov. DOI: [10.1093/nar/gky1049](https://doi.org/10.1093/nar/gky1049). cím: <https://doi.org/10.1093/nar/gky1049> (hivatkozási oldalak 35, 37).
- [117] A. Bateman, M.-J. Martin, S. Orchard, M. Magrane, S. Ahmad, E. Alpi, E. H. Bowler-Barnett, R. Britto, H. Bye-A-Jee, A. Cukura, P. Denny, T. Dogan, T. Ebenezer, J. Fan, P. Garmiri, L. J. da Costa Gonzales, E. Hatton-Ellis, A. Hussein, A. Ignatchenko, G. Insana, R. Ishtiaq, V. Joshi, D. Jyothi, S. Kandasaamy, A. Lock, A. Luciani, M. Lugaric, J. Luo, Y. Lussi, A. MacDougall, F. Madeira, M. Mahmoudy, A. Mishra, K. Moulang, A. Nightingale, S. Pundir, G. Qi, S. Raj, P. Raposo, D. L. Rice, R. Saidi, R. Santos, E. Speretta, J. Stephenson, P. Totoo, E. Turner, N. Tyagi, P. Vasudev, K. Warner, X. Watkins, R. Zaru, H. Zellner, A. J. Bridge, L. Aimò, G. Argoud-Puy, A. H. Auchincloss, K. B. Axelsen, P. Bansal, D. Baratin, T. M. Batista Neto, M.-C. Blatter, J. T. Bolleman, E. Boutet, L. Breuza, B. C. Gil, C. Casals-Casas, K. C. Echioukh, E. Coudert, B. Cuhe, E. de Castro, A. Estreicher, M. L. Famiglietti, M. Feuermann,

- E. Gasteiger, P. Gaudet, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz, C. Hulo, N. Hyka-Nouspikel, F. Junco, A. Kerhornou, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, V. Muthukrishnan, S. Paesano, I. Pedruzzi, S. Pilbout, L. Pourcel, S. Poux, M. Pozzato, M. Pruess, N. Redaschi, C. Rivoire, C. J. A. Sigrist, K. Sonesson, S. Sundaram, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang és J. Zhang, “UniProt: the Universal Protein Knowledgebase in 2023”, *Nucleic Acids Research*, 51. évfolyam, D1. szám, D523–D531. oldal, 2022. nov., ISSN: 1362-4962. DOI: [10.1093/nar/gkac1052](https://doi.org/10.1093/nar/gkac1052). cím: <http://dx.doi.org/10.1093/nar/gkac1052> (hivatkozási oldal 37).
- [118] M. C. Aspromonte, M. V. Nugnes, F. Quaglia, A. Bouharoua, V. Sagris, V. J. Promponas, A. Chasapi, E. Fichó, G. E. Balatti, G. Parisi, M. G. Buitrón, G. Erdos, M. Pajkos, Z. Dosztányi, L. Dobson, A. D. Conte, D. Clementel, E. Salladini, E. Leonardi, F. Kordevani, H. Ghafouri, L. G. T. Ku, A. M. Monzon, C. Ferrari, Z. Kálmán, J. F. Nilsson, J. Santos, C. Pintado-Grima, S. Ventura, V. Ács, R. Pancsa, M. G. Kulik, M. A. Andrade-Navarro, P. J. B. Pereira, S. Longhi, P. L. Mercier, J. Bergier, P. Tompa, T. Lazar, S. C. E. Tosatto és D. Piovesan, “DisProt in 2024: improving function annotation of intrinsically disordered proteins”, *Nucleic Acids Research*, 52. évfolyam, D1. szám, D434–D441. oldal, 2023. okt., ISSN: 1362-4962. DOI: [10.1093/nar/gkad928](https://doi.org/10.1093/nar/gkad928). cím: <http://dx.doi.org/10.1093/nar/gkad928> (hivatkozási oldal 37).
- [119] A. Hatos, B. Hajdu-Soltész, A. M. Monzon, N. Palopoli, L. Álvarez, B. Aykac-Fas, C. Bassot, G. I. Benitez, M. Bevilacqua, A. Chasapi, L. Chemes, N. E. Davey, R. Davidović, A. K. Dunker, A. Elofsson, J. Gobeill, N. S. G. Foutel, G. Sudha, M. Guharoy, T. Horvath, V. Iglesias, A. V. Kajava, O. P. Kovacs, J. Lamb, M. Lambrugh, T. Lazar, J. Y. Leclercq, E. Leonardi, S. Macedo-Ribeiro, M. Macossay-Castillo, E. Maiani, J. A. Manso, C. Marino-Buslje, E. Martinez-Pérez, B. Mészáros, I. Mičetić, G. Minervini, N. Murvai, M. Necci, C. A. Ouzounis, M. Pajkos, L. Paladin, R. Pancsa, E. Papaleo, G. Parisi, E. Pasche, P. J. Barbosa Pereira, V. J. Promponas, J. Pujols, F. Quaglia, P. Ruch, M. Salvatore, E. Schad, B. Szabo, T. Szaniszló, S. Tamana, A. Tantos, N. Veljkovic, S. Ventura, W. Vranken, Z. Dosztányi, P. Tompa, S. C. E. Tosatto és D. Piovesan, “DisProt: intrinsic protein disorder annotation in 2020”, *Nucleic Acids Research*, 2019. nov. DOI: [10.1093/nar/gkz975](https://doi.org/10.1093/nar/gkz975). cím: <https://doi.org/10.1093/nar/gkz975> (hivatkozási oldal 37).
- [120] J. Yang, M. Gao, J. Xiong, Z. Su és Y. Huang, “Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding”, *Protein Science*, 28. évfolyam, 11. szám, 1952–1965. oldal, 2019. szept. DOI: [10.1002/pro.3718](https://doi.org/10.1002/pro.3718). cím: <https://doi.org/10.1002/pro.3718> (hivatkozási oldal 37).
- [121] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn és A. Bateman, “Pfam: The protein families database in 2021”, *Nucleic Acids Research*, 49. évfolyam, D1. szám, D412–D419. oldal, 2020. okt., ISSN: 1362-4962. DOI: [10.1093/nar/gkaa913](https://doi.org/10.1093/nar/gkaa913). cím: <http://dx.doi.org/10.1093/nar/gkaa913> (hivatkozási oldal 37).

- [122] S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto és R. D. Finn, “The Pfam protein families database in 2019”, *Nucleic Acids Research*, 47. évfolyam, D1. szám, D427–D432. oldal, 2018. okt. DOI: [10.1093/nar/gky995](https://doi.org/10.1093/nar/gky995). cím: <https://doi.org/10.1093/nar/gky995> (hivatkozási oldal 39).
- [123] T. G. O. Consortium, “The Gene Ontology Resource: 20 years and still GOing strong”, *Nucleic Acids Res*, 47. évfolyam, D1. szám, D330–D338. oldal, 2019. jan. (hivatkozási oldal 39).
- [124] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin és G. Sherlock, “Gene Ontology: tool for the unification of biology”, *Nature Genetics*, 25. évfolyam, 1. szám, 25–29. oldal, 2000. máj. DOI: [10.1038/75556](https://doi.org/10.1038/75556). cím: <https://doi.org/10.1038/75556> (hivatkozási oldal 39).
- [125] N. Schuurman és A. Leszczynski, “Ontologies for Bioinformatics”, *Bioinformatics and Biology Insights*, 2. évfolyam, BBI.S451., 2008. jan. DOI: [10.4137/bbi.s451](https://doi.org/10.4137/bbi.s451). cím: <https://doi.org/10.4137/bbi.s451> (hivatkozási oldal 39).
- [126] H. M. Berman, “The Protein Data Bank”, *Nucleic Acids Research*, 28. évfolyam, 1. szám, 235–242. oldal, 2000. jan. DOI: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235). cím: <https://doi.org/10.1093/nar/28.1.235> (hivatkozási oldal 39).
- [127] S. K. Burley, H. M. Berman, C. Bhikadiya, C. Bi, L. Chen, L. Di Costanzo, C. Christie, K. Dalenberg, J. M. Duarte, S. Dutta, Z. Feng, S. Ghosh, D. S. Goodsell, R. K. Green, V. Guranović, D. Guzenko, B. P. Hudson, T. Kalro, Y. Liang, R. Lowe, H. Namkoong, E. Peisach, I. Periskova, A. Prlić, C. Randle, A. Rose, P. Rose, R. Sala, M. Sekharan, C. Shao, L. Tan, Y.-P. Tao, Y. Valasatava, M. Voigt, J. Westbrook, J. Woo, H. Yang, J. Young, M. Zhuravleva és C. Zardecki, “RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy”, *Nucleic Acids Research*, 47. évfolyam, D1. szám, D464–D474. oldal, 2018. okt. DOI: [10.1093/nar/gky1004](https://doi.org/10.1093/nar/gky1004). cím: <https://doi.org/10.1093/nar/gky1004> (hivatkozási oldal 40).
- [128] V. P. Waman, N. Bordin, R. Alcraft, R. Vickerstaff, C. Rauer, Q. Chan, I. Sillitoe, H. Yamamori és C. Orengo, “CATH 2024: CATH-AlphaFlow Doubles the Number of Structures in CATH and Reveals Nearly 200 New Folds”, *Journal of Molecular Biology*, 168551. oldal, 2024. márc., ISSN: 0022-2836. DOI: [10.1016/j.jmb.2024.168551](https://doi.org/10.1016/j.jmb.2024.168551). cím: <http://dx.doi.org/10.1016/j.jmb.2024.168551> (hivatkozási oldal 40).
- [129] I. Sillitoe, N. Dawson, T. E. Lewis, S. Das, J. G. Lees, P. Ashford, A. Tolulope, H. M. Scholes, I. Senatorov, A. Bujan, F. Ceballos Rodriguez-Conde, B. Dowling, J. Thornton és C. A. Orengo, “CATH: expanding the horizons of structure-based functional annotations for genome sequences”, *Nucleic Acids Research*, 47. évfolyam, D1. szám, D280–D284. oldal, 2018. nov. DOI: [10.1093/nar/gky1097](https://doi.org/10.1093/nar/gky1097). cím: <https://doi.org/10.1093/nar/gky1097> (hivatkozási oldal 40).

- [130] S. M. Sunkin, L. Ng, C. Lau, T. Dolbeare, T. L. Gilbert, C. L. Thompson, M. Hawrylycz és C. Dang, “Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system”, *Nucleic Acids Research*, 41. évfolyam, D1. szám, D996–D1008. oldal, 2012. nov. DOI: [10.1093/nar/gks1042](https://doi.org/10.1093/nar/gks1042). cím: <https://doi.org/10.1093/nar/gks1042> (hivatkozási oldal 41).
- [131] Z. E. Kalman, D. Dudola, B. Mészáros, Z. Gáspári és L. Dobson, “PSINDB: the postsynaptic protein–protein interaction database”, *Database*, 2022. évfolyam, 2022. jan. DOI: [10.1093/database/baac007](https://doi.org/10.1093/database/baac007). cím: <https://doi.org/10.1093/database/baac007> (hivatkozási oldal 41).
- [132] F. Pucci, J. M. Kwasigroch és M. Rooman, “Protein Thermal Stability Engineering Using HoTMuSiC”, *Methods in Molecular Biology*, Springer US, 2020, 59–73. oldal. DOI: [10.1007/978-1-0716-0270-6_5](https://doi.org/10.1007/978-1-0716-0270-6_5). cím: https://doi.org/10.1007/978-1-0716-0270-6_5 (hivatkozási oldal 41).
- [133] J. Bendl, J. Stourac, O. Salanda, A. Pavelka, E. D. Wieben, J. Zendulka, J. Brezovsky és J. Damborsky, “PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations”, *PLoS Computational Biology*, 10. évfolyam, 1. szám, P. P. Gardner, szerkesztő, e1003440., 2014. jan. DOI: [10.1371/journal.pcbi.1003440](https://doi.org/10.1371/journal.pcbi.1003440). cím: <https://doi.org/10.1371/journal.pcbi.1003440> (hivatkozási oldal 42).
- [134] H. M. Dingerdissen, J. Torcivia-Rodriguez, Y. Hu, T.-C. Chang, R. Mazumder és R. Kahsay, “BioMuta and BioXpress: mutation and expression knowledgebases for cancer biomarker discovery”, *Nucleic Acids Research*, 46. évfolyam, D1. szám, D1128–D1136. oldal, 2017. okt. DOI: [10.1093/nar/gkx907](https://doi.org/10.1093/nar/gkx907). cím: <https://doi.org/10.1093/nar/gkx907> (hivatkozási oldal 42).
- [135] K. Olechnovič és Č. Venclovas, “Contact Area-Based Structural Analysis of Proteins and Their Complexes Using CAD-Score”, *Methods in Molecular Biology*, Springer US, 2020, 75–90. oldal. DOI: [10.1007/978-1-0716-0270-6_6](https://doi.org/10.1007/978-1-0716-0270-6_6). cím: https://doi.org/10.1007/978-1-0716-0270-6_6 (hivatkozási oldal 42).
- [136] —, “The CAD-score web server: contact area-based comparison of structures and interfaces of proteins, nucleic acids and their complexes”, *Nucleic Acids Research*, 42. évfolyam, W1. szám, W259–W263., 2014. máj. DOI: [10.1093/nar/gku294](https://doi.org/10.1093/nar/gku294). cím: <https://doi.org/10.1093/nar/gku294> (hivatkozási oldal 42).
- [137] M. Giollo, A. J. Martin, I. Walsh, C. Ferrari és S. C. Tosatto, “NeEMO: a method using residue interaction networks to improve prediction of protein stability upon mutation”, *BMC Genomics*, 15. évfolyam, S4. szám, 2014. máj. DOI: [10.1186/1471-2164-15-s4-s7](https://doi.org/10.1186/1471-2164-15-s4-s7). cím: <https://doi.org/10.1186/1471-2164-15-s4-s7> (hivatkozási oldalak 43, 63, 64).
- [138] R. Erban, S. J. Chapman, Philip és K. Maini, *A practical guide to stochastic simulations of reaction-diffusion processes*, 35 pages, available as <http://arxiv.org/abs/0704.1908>, 2007 (hivatkozási oldal 44).

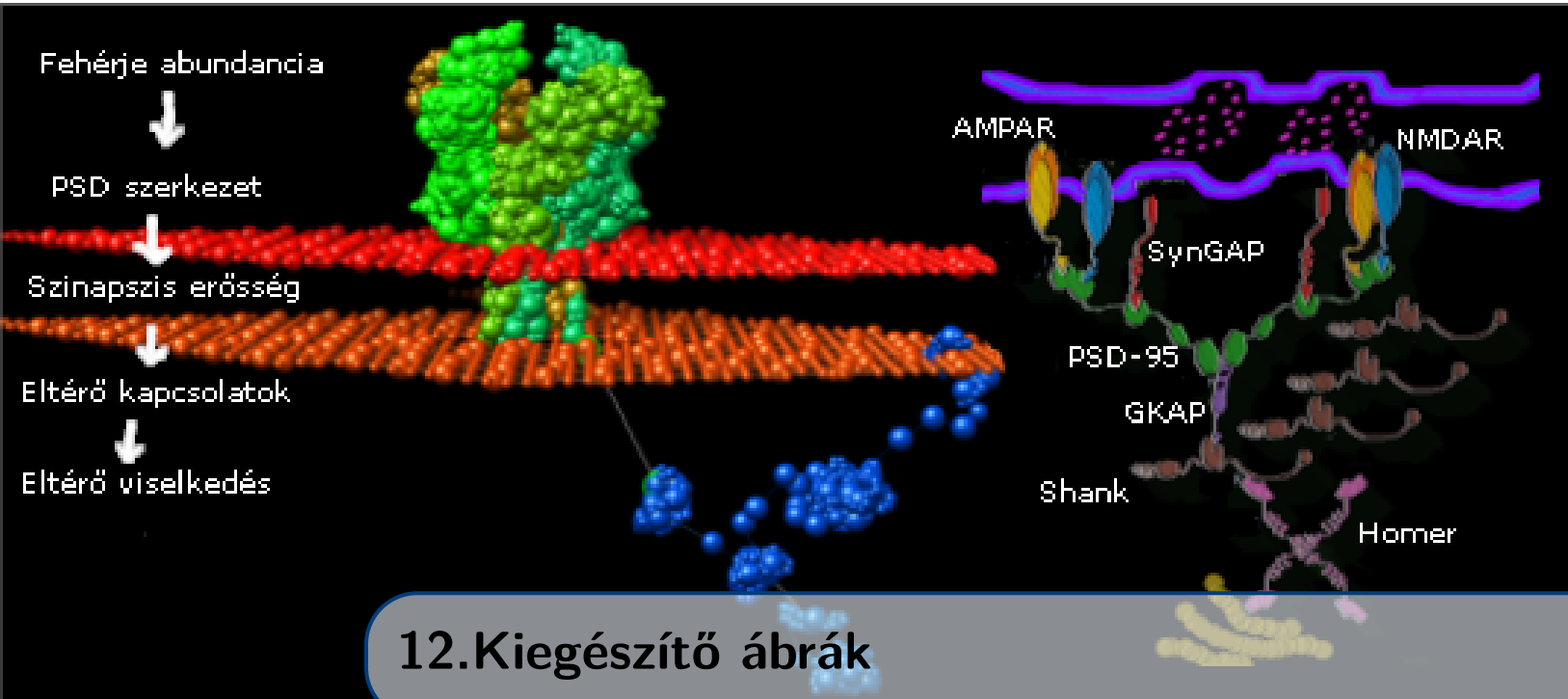
-
- [139] S. Rizzetto, P. Moyses, B. Baldacci, C. Priami és A. Csikász-Nagy, “Context-dependent prediction of protein complexes by SiComPre”, *npj Systems Biology and Applications*, 4. évfolyam, 1. szám, 2018. szept. DOI: [10.1038/s41540-018-0073-0](https://doi.org/10.1038/s41540-018-0073-0). cím: <https://doi.org/10.1038/s41540-018-0073-0> (hivatkozási oldal 45).
- [140] C. M., *Pattern Recognition and Machine Learning*, en, 1. kiadás, Information Science and Statistics sorozat. New York, NY: Springer, 2006. aug. (hivatkozási oldal 45).
- [141] J. A. Hartigan és M. A. Wong, “Algorithm AS 136: A K-Means Clustering Algorithm”, *Applied Statistics*, 28. évfolyam, 1. szám, 100. oldal, 1979. DOI: [10.2307/2346830](https://doi.org/10.2307/2346830). cím: <https://doi.org/10.2307/2346830> (hivatkozási oldalak 45, 68).
- [142] D. T. Pham, S. S. Dimov és C. D. Nguyen, “Selection of iK/i in iK/i-means clustering”, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219. évfolyam, 1. szám, 103–119. oldal, 2005. jan. DOI: [10.1243/095440605x8298](https://doi.org/10.1243/095440605x8298). cím: <https://doi.org/10.1243/095440605x8298> (hivatkozási oldal 46).
- [143] K. R. Shahapure és C. Nicholas, “Cluster Quality Analysis Using Silhouette Score”, *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, 2020. okt. DOI: [10.1109/dsaa49011.2020.00096](https://doi.org/10.1109/dsaa49011.2020.00096). cím: <https://doi.org/10.1109/dsaa49011.2020.00096> (hivatkozási oldal 46).
- [144] S. Wold, K. Esbensen és P. Geladi, “Principal component analysis”, *Chemometrics and Intelligent Laboratory Systems*, 2. évfolyam, 1-3. szám, 37–52. oldal, 1987. aug. DOI: [10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9). cím: [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9) (hivatkozási oldal 47).
- [145] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt és SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”, *Nature Methods*, 17. évfolyam, 261–272. oldal, 2020. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2) (hivatkozási oldal 48).
- [146] G. Li, M. Taljaard, E. R. Van den Heuvel, M. A. Levine, D. J. Cook, G. A. Wells, P. J. Devereaux és L. Thabane, “An introduction to multiplicity issues in clinical trials: the what, why, when and how”, *International Journal of Epidemiology*, dyw320., 2016. dec., ISSN: 1464-3685. DOI: [10.1093/ije/dyw320](https://doi.org/10.1093/ije/dyw320). cím: <http://dx.doi.org/10.1093/ije/dyw320> (hivatkozási oldal 51).
- [147] T. V. Perneger, “What’s wrong with Bonferroni adjustments”, *BMJ*, 316. évfolyam, 7139. szám, 1236–1238. oldal, 1998. ápr., ISSN: 1468-5833. DOI: [10.1136/bmj.316.7139.1236](https://doi.org/10.1136/bmj.316.7139.1236). cím: <http://dx.doi.org/10.1136/bmj.316.7139.1236> (hivatkozási oldal 52).
- [148] W. Haynes, “Benjamini–Hochberg Method”, *Encyclopedia of Systems Biology*, W. Dubitzky, O. Wolkenhauer, K.-H. Cho és H. Yokota, szerkesztők. New York, NY: Springer New York, 2013, 78–78. oldal, ISBN: 978-1-4419-9863-7. DOI: [10.1007/978-1-](https://doi.org/10.1007/978-1-4419-9863-7)

- 4419-9863-7_1215. cím: https://doi.org/10.1007/978-1-4419-9863-7_1215 (hivatkozási oldal 52).
- [149] J. Smith és J. Doe, “Understanding AMPAR Structure and Function”, *Journal of Neuroscience*, 40. évfolyam, 123–134. oldal, 2020. DOI: [10.1234/j.neuro.2020.01.001](https://doi.org/10.1234/j.neuro.2020.01.001) (hivatkozási oldal 52).
- [150] A. Brown és R. Green, “The Role of AMPAR in Synaptic Plasticity”, *Nature Neuroscience*, 21. évfolyam, 567–576. oldal, 2018. DOI: [10.1038/s41593-018-0123-4](https://doi.org/10.1038/s41593-018-0123-4) (hivatkozási oldalak 52, 53).
- [151] E. Johnson és M. Lee, “AMPA Complexes and Their Regulation”, *PLoS Biology*, 19. évfolyam, e3001254., 2021. DOI: [10.1371/journal.pbio.3001254](https://doi.org/10.1371/journal.pbio.3001254) (hivatkozási oldalak 52, 53).
- [152] M. L. Mayer és N. Armstrong, *Crystal structure of the GluA2 ligand-binding core in complex with kainate*, Protein Data Bank, PDB ID: 3KG2, 2009 (hivatkozási oldal 53).
- [153] E. C. Twomey és A. I. Sobolevsky, *Structural Mechanisms of Gating in the GluA2 AMPA Receptor*, Protein Data Bank, PDB ID: 5WEO, 2017 (hivatkozási oldal 53).
- [154] G. A. Tavares, E. H. Panepucci és A. T. Brunger, “Structural Characterization of the Intramolecular Interaction between the SH3 and Guanylate Kinase Domains of PSD-95”, *Molecular Cell*, 8. évfolyam, 6. szám, 1313–1325. oldal, 2001. dec. DOI: [10.1016/s1097-2765\(01\)00416-6](https://doi.org/10.1016/s1097-2765(01)00416-6). cím: [https://doi.org/10.1016/s1097-2765\(01\)00416-6](https://doi.org/10.1016/s1097-2765(01)00416-6) (hivatkozási oldal 53).
- [155] Z. Harmat, D. Dudola és Z. Gáspári, “DIPEND: An Open-Source Pipeline to Generate Ensembles of Disordered Segments Using Neighbor-Dependent Backbone Preferences”, *Biomolecules*, 11. évfolyam, 10. szám, 1505. oldal, 2021. okt. DOI: [10.3390/biom11101505](https://doi.org/10.3390/biom11101505). cím: <https://doi.org/10.3390/biom11101505> (hivatkozási oldal 54).
- [156] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Židek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis és S. Velankar, “AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models”, *Nucleic Acids Research*, 50. évfolyam, D1. szám, D439–D444. oldal, 2021. nov. DOI: [10.1093/nar/gkab1061](https://doi.org/10.1093/nar/gkab1061). cím: <https://doi.org/10.1093/nar/gkab1061> (hivatkozási oldal 54).
- [157] L. Schermelleh, A. Ferrand, T. Huser, C. Eggeling, M. Sauer, O. Biehlmaier és G. P. C. Drummen, “Super-resolution microscopy demystified”, *Nature Cell Biology*, 21. évfolyam, 1. szám, 72–84. oldal, 2019. jan. DOI: [10.1038/s41556-018-0251-8](https://doi.org/10.1038/s41556-018-0251-8). cím: <https://doi.org/10.1038/s41556-018-0251-8> (hivatkozási oldal 55).
- [158] D. Cheng, C. C. Hoogenraad, J. Rush, E. Ramm, M. A. Schlager, D. M. Duong, P. Xu, S. R. Wijayawardana, J. Hanfelt, T. Nakagawa, M. Sheng és J. Peng, “Relative and Absolute Quantification of Postsynaptic Density Proteome Isolated from Rat Forebrain and Cerebellum”, *Molecular & Cellular Proteomics*, 5. évfolyam, 6. szám,

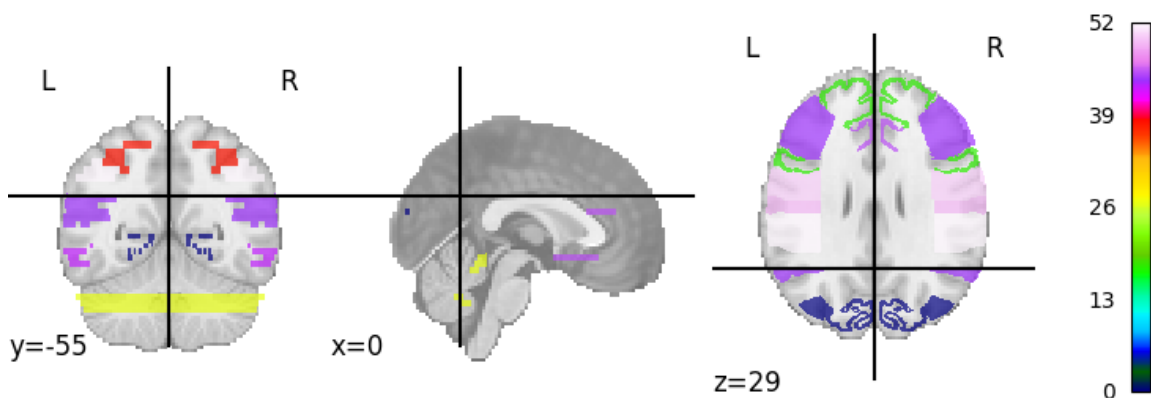
- 1158–1170. oldal, 2006. jún. DOI: [10.1074/mcp.d500009-mcp200](https://doi.org/10.1074/mcp.d500009-mcp200). cím: <https://doi.org/10.1074/mcp.d500009-mcp200> (hivatkozási oldal 56).
- [159] M. Farley, “Structure and Composition of Postsynaptic Densities”, *The University of Texas MD Anderson Cancer Center UTHHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*, 2015. aug. cím: https://digitalcommons.library.tmc.edu/utgsbs_dissertations/603 (hivatkozási oldal 56).
- [160] A. Sali, R. Glaeser, T. Earnest és W. Baumeister, “From words to literature in structural proteomics”, *Nature*, 422. évfolyam, 6928. szám, 216–225. oldal, 2003. márc. DOI: [10.1038/nature01513](https://doi.org/10.1038/nature01513). cím: <https://doi.org/10.1038/nature01513> (hivatkozási oldal 57).
- [161] S. Grillner, A. Kozlov és J. H. Kotaleski, “Integrative neuroscience: linking levels of analyses”, *Current Opinion in Neurobiology*, 15. évfolyam, 5. szám, 614–621. oldal, 2005. okt. DOI: [10.1016/j.conb.2005.08.017](https://doi.org/10.1016/j.conb.2005.08.017). cím: <https://doi.org/10.1016/j.conb.2005.08.017> (hivatkozási oldal 57).
- [162] T. C. Terwilliger, P. D. Adams, P. V. Afonine és O. V. Sobolev, “Cryo-EM map interpretation and protein model-building using iterative map segmentation”, *Protein Science*, 29. évfolyam, 1. szám, 87–99. oldal, 2019. okt. DOI: [10.1002/pro.3740](https://doi.org/10.1002/pro.3740). cím: <https://doi.org/10.1002/pro.3740> (hivatkozási oldal 58).
- [163] J. He, P. Lin, J. Chen, H. Cao és S.-Y. Huang, “Model building of protein complexes from intermediate-resolution cryo-EM maps with deep learning-guided automatic assembly”, *Nature Communications*, 13. évfolyam, 1. szám, 2022. júl. DOI: [10.1038/s41467-022-31748-9](https://doi.org/10.1038/s41467-022-31748-9). cím: <https://doi.org/10.1038/s41467-022-31748-9> (hivatkozási oldal 58).
- [164] J. H. Jung, X. Chen és T. S. Reese, “Cryo-EM tomography and automatic segmentation delineate modular structures in the postsynaptic density”, *Frontiers in Synaptic Neuroscience*, 15. évfolyam, 2023, ISSN: 1663-3563. DOI: [10.3389/fnsyn.2023.1123564](https://doi.org/10.3389/fnsyn.2023.1123564). cím: <https://www.frontiersin.org/articles/10.3389/fnsyn.2023.1123564> (hivatkozási oldal 58).
- [165] D. T. Gillespie, “Stochastic Simulation of Chemical Kinetics”, *Annual Review of Physical Chemistry*, 58. évfolyam, 1. szám, 35–55. oldal, 2007. máj. DOI: [10.1146/annurev.physchem.58.032806.104637](https://doi.org/10.1146/annurev.physchem.58.032806.104637). cím: <https://doi.org/10.1146/annurev.physchem.58.032806.104637> (hivatkozási oldal 61).
- [166] J. G. Tate, S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jupe, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell és S. A. Forbes, “COSMIC: the Catalogue Of Somatic Mutations In Cancer”, *Nucleic Acids Research*, 47. évfolyam, D1. szám, D941–D947. oldal, 2018. okt. DOI: [10.1093/nar/gky1015](https://doi.org/10.1093/nar/gky1015). cím: <https://doi.org/10.1093/nar/gky1015> (hivatkozási oldal 64).
- [167] H. Li, Q. Xia és Y. Wang, “Research and Improvement of Kruskal Algorithm”, *Journal of Computer and Communications*, 5. évfolyam, 12. szám, 63–69. oldal, 2017. DOI: [10.4236/jcc.2017.512007](https://doi.org/10.4236/jcc.2017.512007). cím: <https://doi.org/10.4236/jcc.2017.512007> (hivatkozási oldal 65).

- [168] W. G. Walkup, T. L. Mastro, L. T. Schenker, J. Vielmetter, R. Hu, A. Iancu, M. Reghunathan, B. D. Bannon és M. B. Kennedy, “A model for regulation by SynGAP-1 of binding of synaptic proteins to PDZ-domain ‘Slots’ in the postsynaptic density”, *eLife*, 5. évfolyam, 2016. szept. DOI: [10.7554/elife.16813](https://doi.org/10.7554/elife.16813). cím: <https://doi.org/10.7554/elife.16813> (hivatkozási oldal 72).
- [169] M. Miski, Á. Weber, K. Fekete-Molnár, B. M. Keömley-Horváth, A. Csikász-Nagy és Z. Gáspári, “Simulated complexes formed from a set of postsynaptic proteins suggest a localised effect of a hypomorphic Shank mutation”, *BMC Neuroscience*, 25. évfolyam, 1. szám, 2024. júl., ISSN: 1471-2202. DOI: [10.1186/s12868-024-00880-1](http://dx.doi.org/10.1186/s12868-024-00880-1). cím: <http://dx.doi.org/10.1186/s12868-024-00880-1> (hivatkozási oldalak 76, 93).
- [170] M. K. Hayashi, H. M. Ames és Y. Hayashi, “Tetrameric Hub Structure of Postsynaptic Scaffolding Protein Homer”, *The Journal of Neuroscience*, 26. évfolyam, 33. szám, 8492–8501. oldal, 2006. aug. DOI: [10.1523/jneurosci.2731-06.2006](https://doi.org/10.1523/jneurosci.2731-06.2006). cím: <https://doi.org/10.1523/jneurosci.2731-06.2006> (hivatkozási oldal 77).
- [171] M. Miski, B. M. Keömley-Horváth, D. R. Megyeriné, A. Csikász-Nagy és Z. Gáspári, “Diversity of synaptic protein complexes as a function of the abundance of their constituent proteins: A modeling approach”, *PLOS Computational Biology*, 18. évfolyam, 1. szám, M. Migliore, szerkesztő, e1009758., 2022. jan. DOI: [10.1371/journal.pcbi.1009758](https://doi.org/10.1371/journal.pcbi.1009758). cím: <https://doi.org/10.1371/journal.pcbi.1009758> (hivatkozási oldalak 81, 82, 91, 92).
- [172] A. Sudarov, “Defining the Role of Cerebellar Purkinje Cells in Autism Spectrum Disorders”, *The Cerebellum*, 12. évfolyam, 6. szám, 950–955. oldal, 2013. máj. DOI: [10.1007/s12311-013-0490-y](https://doi.org/10.1007/s12311-013-0490-y). cím: <https://doi.org/10.1007/s12311-013-0490-y> (hivatkozási oldalak 84, 93).
- [173] H. L. Wilson, “Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms”, *Journal of Medical Genetics*, 40. évfolyam, 8. szám, 575–584. oldal, 2003. aug. DOI: [10.1136/jmg.40.8.575](https://doi.org/10.1136/jmg.40.8.575). cím: <https://doi.org/10.1136/jmg.40.8.575> (hivatkozási oldalak 85, 93).
- [174] P. T. Tsai, “Autism and cerebellar dysfunction: Evidence from animal models”, *Seminars in Fetal and Neonatal Medicine*, 21. évfolyam, 5. szám, 349–355. oldal, 2016. okt. DOI: [10.1016/j.siny.2016.04.009](https://doi.org/10.1016/j.siny.2016.04.009). cím: <https://doi.org/10.1016/j.siny.2016.04.009> (hivatkozási oldalak 85, 93).
- [175] M. Miski, K. Kornél, A. Csikász-Nagy és Z. Gáspári, “Integrative modeling of possible layouts of Kir2.1 localized in membranes and connected by PSD-95”, *JEDLIK LABORATORIES REPORTS*, VII. évfolyam, 29–32. oldal, 2019, ISSN: 2064-3942 (hivatkozási oldalak 85, 94).
- [176] M. T. Muhammed és E. Aki-Yalcin, “Homology modeling in drug discovery: Overview, current applications, and future perspectives”, *Chemical Biology & Drug Design*, 93. évfolyam, 1. szám, 12–20. oldal, 2018. okt. DOI: [10.1111/cbdd.13388](https://doi.org/10.1111/cbdd.13388). cím: <https://doi.org/10.1111/cbdd.13388> (hivatkozási oldal 86).

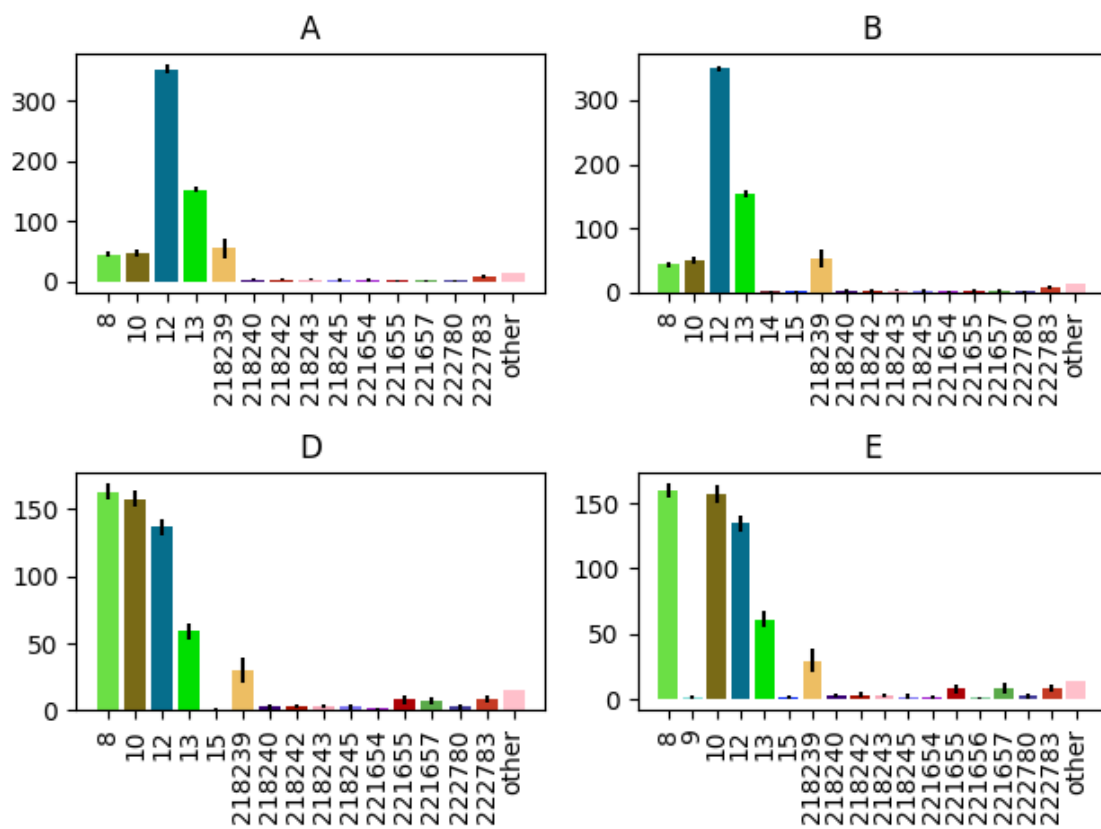
-
- [177] A. Kiss-Tóth, L. Dobson, B. Péterfia, A. F. Ángyán, B. Ligeti, G. Lukács és Z. Gáspári, “Occurrence of Ordered and Disordered Structural Elements in Postsynaptic Proteins Supports Optimization for Interaction Diversity”, *Entropy*, 21. évfolyam, 8. szám, 761. oldal, 2019. aug. DOI: [10.3390/e21080761](https://doi.org/10.3390/e21080761). cím: <https://doi.org/10.3390/e21080761> (hivatkozási oldal 88).
- [178] A. Ö. Sungur, T. M. Redecker, E. Andres, W. Dürichen, R. K. W. Schwarting, A. del Rey és M. Wöhr, “Reduced Efficacy of d-Amphetamine and 3, 4-Methylenedioxymethamphetamine in Inducing Hyperactivity in Mice Lacking the Postsynaptic Scaffolding Protein SHANK1”, *Frontiers in Molecular Neuroscience*, 11. évfolyam, 2018. nov. DOI: [10.3389/fnmol.2018.00419](https://doi.org/10.3389/fnmol.2018.00419). cím: <https://doi.org/10.3389/fnmol.2018.00419> (hivatkozási oldal 89).
- [179] D. Pascovici, J. X. Wu, M. J. McKay, C. Joseph, Z. Noor, K. Kamath, Y. Wu, S. Ranganathan, V. Gupta és M. Mirzaei, “Clinically Relevant Post-Translational Modification Analyses—Maturing Workflows and Bioinformatics Tools”, *International Journal of Molecular Sciences*, 20. évfolyam, 1. szám, 16. oldal, 2018. dec. DOI: [10.3390/ijms20010016](https://doi.org/10.3390/ijms20010016). cím: <https://doi.org/10.3390/ijms20010016> (hivatkozási oldal 90).
- [180] V. A. Petyuk, W.-J. Qian, M. H. Chin, H. Wang, E. A. Livesay, M. E. Monroe, J. N. Adkins, N. Jaitly, D. J. Anderson, D. G. Camp, D. J. Smith és R. D. Smith, “Spatial mapping of protein abundances in the mouse brain by voxelation integrated with high-throughput liquid chromatography-mass spectrometry”, *Genome Research*, 17. évfolyam, 3. szám, 328–336. oldal, 2007. febr. DOI: [10.1101/gr.5799207](https://doi.org/10.1101/gr.5799207). cím: <https://doi.org/10.1101/gr.5799207> (hivatkozási oldal 91).
- [181] M. Miski és A. Csikász-Nagy, “Analyses of protein-protein interactions in the PSD by stochastic simulations”, *PHD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY AND BIONICS*, 15. évfolyam, 88–90. oldal, 2020, ISSN: 2064-7271 (hivatkozási oldal 92).
- [182] C. E. Holt, K. C. Martin és E. M. Schuman, “Local translation in neurons: visualization and function”, *Nature Structural & Molecular Biology*, 26. évfolyam, 7. szám, 557–566. oldal, 2019. júl. DOI: [10.1038/s41594-019-0263-5](https://doi.org/10.1038/s41594-019-0263-5). cím: <https://doi.org/10.1038/s41594-019-0263-5> (hivatkozási oldal 93).
- [183] M. Miski, “Analyses of protein-protein interactions related to nNOS by stochastic simulations”, *PHD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY AND BIONICS*, vol. 14 évfolyam, 59–62. oldal, 2019, ISSN: 2064-7271 (hivatkozási oldal 94).
- [184] M. Miski és A. Csikász-Nagy, “Distribution and structure of postsynaptic protein complexes assessed by simulations”, *PHD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY AND BIONICS*, vol. 16 évfolyam, 50. oldal, 2021, ISSN: 2064-7271 (hivatkozási oldal 94).



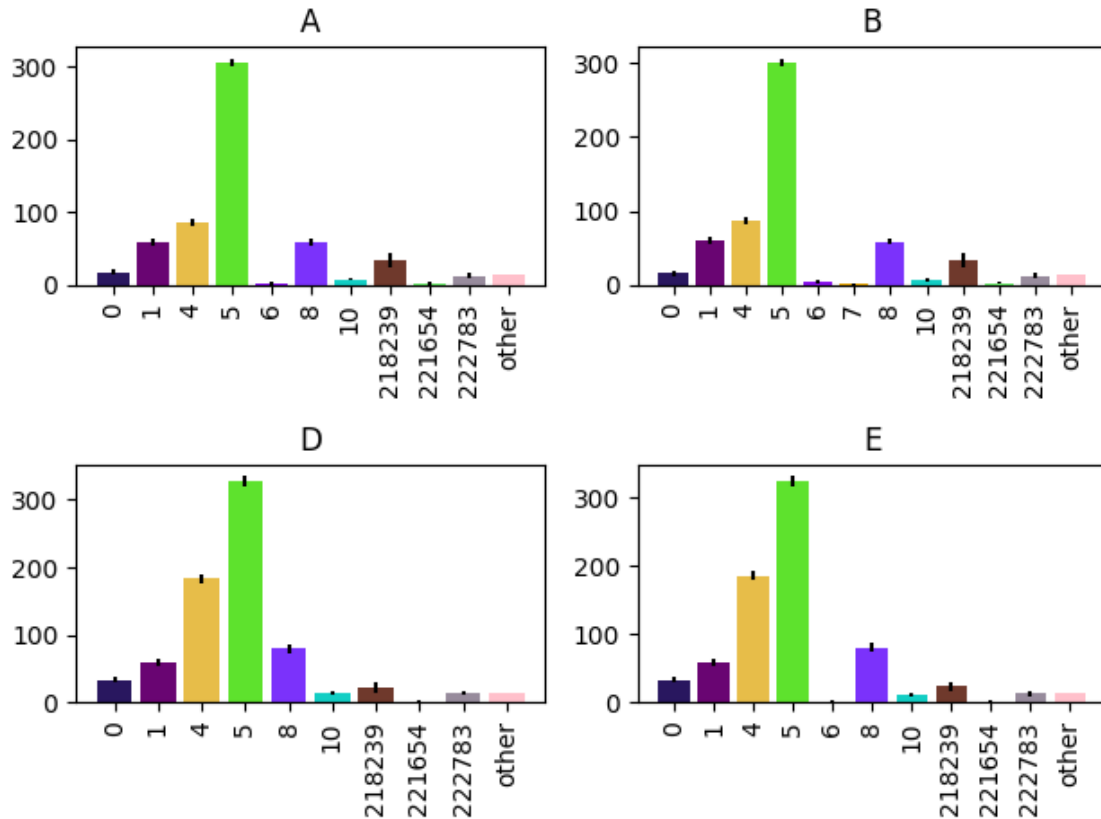
12.Kiegészítő ábrák



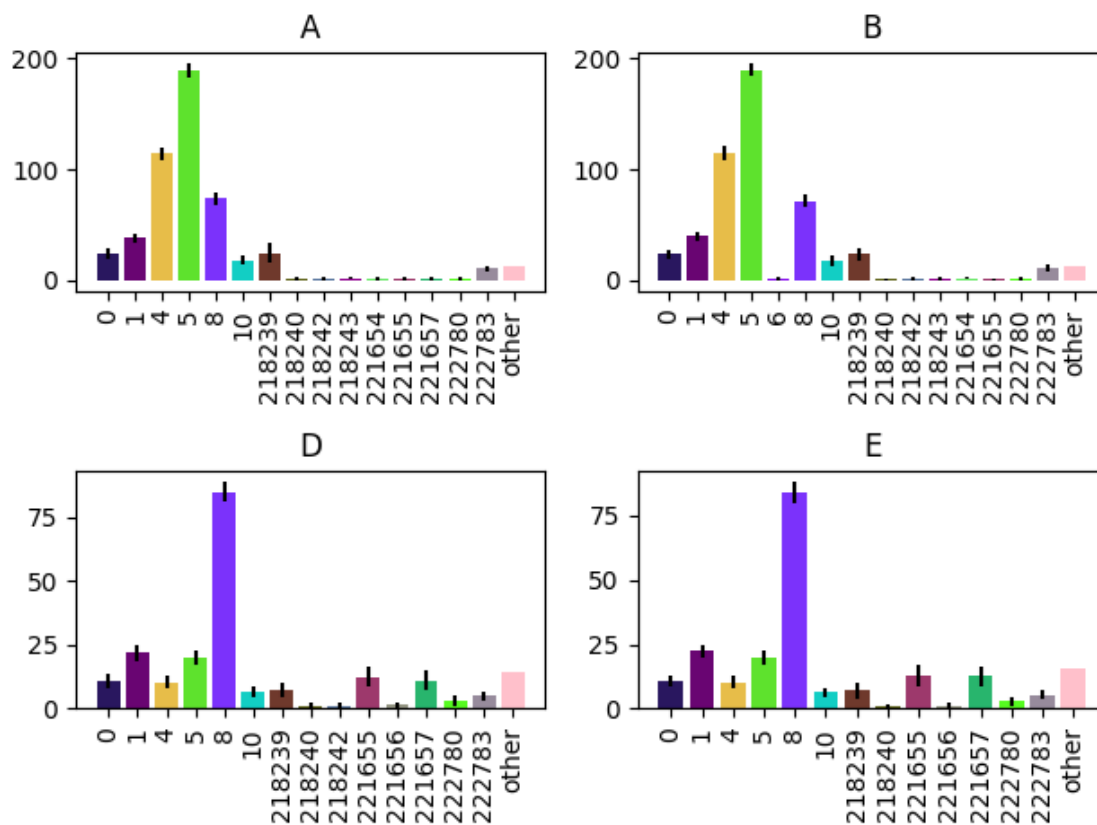
S1. ábra. **p-értékek agyra vetítve** Az ábra három nézetben mutatja az agyat, amelyeken különböző színek jelzik a p-értékek eloszlását az agyterületeken. A színek intenzitása a p-értékek nagyságát jelzi, a színskála pedig 0-tól 52-ig terjedő értékeket ábrázol. Mivel a p-értékek alacsonyka önmagában, az adott értelmezhető és színezhető intervallumra vetítettem. Magas szignifikancia (alacsony p-érték), ami azt jelzi, hogy az adott területen megfigyelt változások nagy valószínűséggel nem véletlenek. Az ábra célja illusztráció.



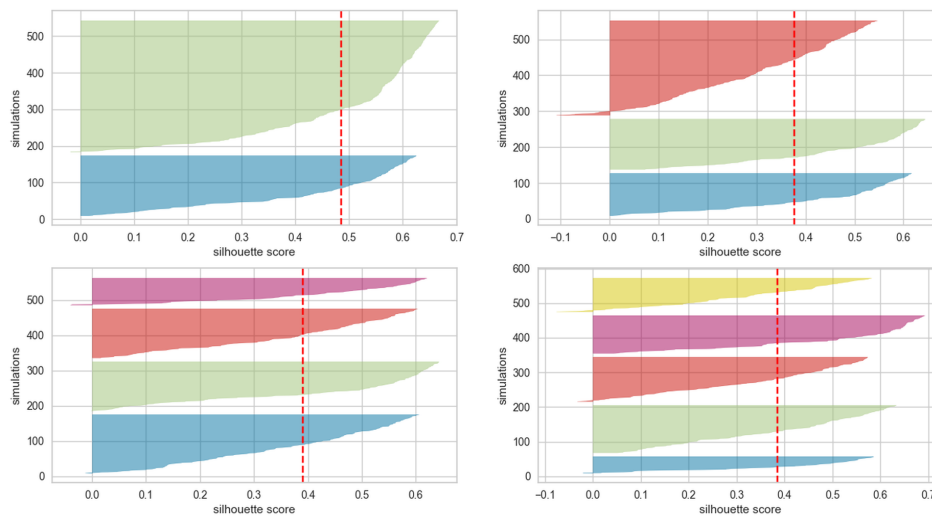
S2. ábra. **Komplex abundanciák a régiókban:** A) H376.XI.50_CBC vad típus B) H376.XI.50_CBC mutáns C) H376.XI.52_CBC vad típus D) H376.XI.52_CBC mutáns.



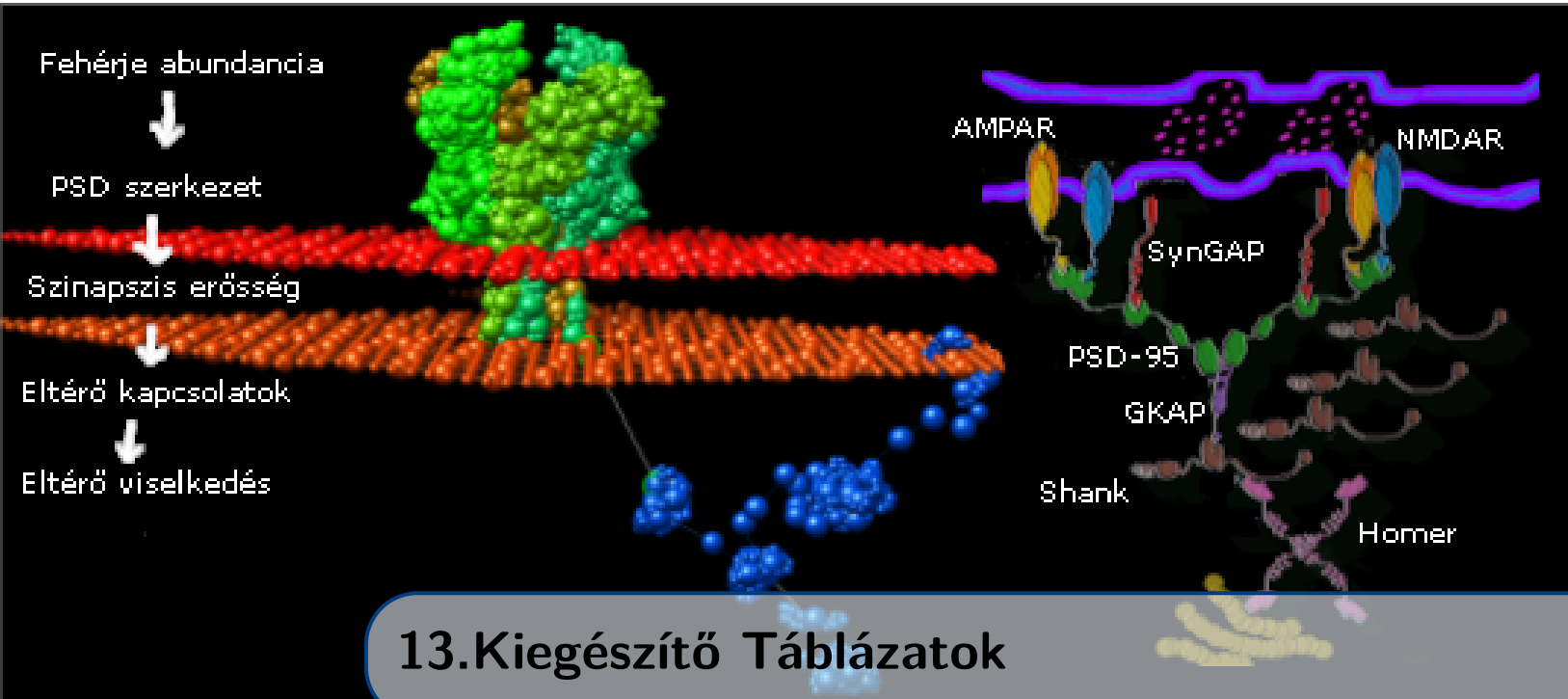
S3. ábra. **Komplex abundanciák a régiókban::** A) H376.IIIB.53_M1C-S1C vadttípus B) H376.IIIB.53_M1C-S1C mutáns D) H376.IV.54_STR vadttípus E) H376.IV.54_STR mutáns. A komplexek a következők: 0:NMDAR/PSD-95;1:NMDAR/PSD-95/SYNGAP;4:PSD-95/AMPAR;5:PSD-95/AMPAR/SYNGAP;6:PSD-95/AMPAR/SYNGAP/GKAP;8:PSD-95/SYNGAP;10:PSD-95/GKAP;218239:GKAP/SHank1;221654:GKAP/SHank1/Homer1;222783:Homer1-tetramer. A ábra (Wild-type): A wild-type mintában a legnagyobb mennyiségben a PSD-95/AMPAR/SYNGAP komplex (zöld) található, amit a PSD-95/AMPAR (sárga) és a NMDAR/PSD-95/SYNGAP (kék) komplexek követnek. B ábra (Mutáns): A mutáns mintában hasonló eloszlás figyelhető meg, de bizonyos komplexek mennyisége csökkent, jelezve, hogy a mutáció hatással van a komplexek kialakulására. D ábra (STR vadttípus és mutáns): Az STR vadttípus és mutáns minták összehasonlítása hasonló trendeket mutat, de a mutáns minta esetében ismét csökkenés figyelhető meg egyes komplexek mennyiségében.



S4. ábra. **Komplex abundanciák a régiókban:** A) H376.XI.50_HIP vad típus B) H376.XI.50_HIP mutáns C) H376.VIII.53_MD vad típus D) H376.VIII.53_MD mutáns. A komplexek a következők: 0:NMDAR/PSD-95;1:NMDAR/PSD-95/SYNGAP; 4:PSD-95/AMPA;5:PSD-95/AMPA/SYNGAP;6:PSD-95/AMPA/SYNGAP/GKAP;8:PSD-95/SYNGAP;10:PSD-95/GKAP;218239:GKAP/SHank1;218240;218242 és 218243: variációi a GKAP/SHank1/SHank1;221654:GKAP/SHank1/Homer1;222783:Homer1-tetramer



S5. ábra. Szilouette értékek különböző mennyiségű klaszter esetén. (K-means). **K=2 Klaszter:** A bal felső kép azt mutatja, hogy két klaszter esetén az adatpontok nagy része pozitív szilouett értékkel rendelkezik, ami azt jelzi, hogy a klaszterezés viszonylag jó. **K=3 Klaszter:** A jobb felső kép azt mutatja, hogy három klaszter esetén is sok adatpont pozitív szilouett értékkel rendelkezik, de néhány negatív érték is megjelenik, ami azt jelzi, hogy néhány adatpont nem illeszkedik jól a klaszterekhez. **K=4 Klaszter:** A bal alsó kép négy klaszter esetén mutatja az eloszlást, ahol az adatpontok többsége még mindig pozitív szilouett értékkel rendelkezik, de a negatív értékek aránya növekszik. **K=5 Klaszter:** A jobb alsó kép azt mutatja, hogy öt klaszter esetén az adatpontok között több a negatív szilouett érték, ami arra utal, hogy az öt klaszter nem feltétlenül optimális az adatok klaszterezéséhez.



13.Kiegészítő Táblázatok

virtual domains are non-globular binding regions defined to represent the interactions
interactions are defined with proteinname_domainname of the two interacting proteins

Input Proteins with Uniprot ID and Domains			
Protein Name	Gene Symbol	UniProt ID	Domains (or virtual domains)
NMDAR	GRIN2B	Q05586	"CNMDAR"
AMPA	GRIA1	P42261	"CAMPA"
PSD-95	DLG4	P78352	"PDZ","PDZ","PDZ","SH3","GK"
SynGAP	SynGAP	Q96PV0	"CSYNGAP"
GKAP	DLGAP1	Q5VSY0	"GBRS","GH","CGKAP"
Shank1	Shank1	Q9Y566	"SH3","PDZ","SAM","virtualEVH1"
Homer1	homer1	Q86YM7	"EVH1"

Domain-domain interactions	
Domain1	Domain2
HOMER1_EVH1	SHANK1_virtualEVH1
HOMER1_HK	HOMER1_HK
GKAP1_CGKAP	SHANK1_PDZ
DLG4_PDZ	GRIN2B_CNMDAR
DLG4_PDZ	GRIA1_CAMPA
DLG4_PDZ	SYNGAP1_CSYNGAP
DLG4_GK	GKAP1_GBRS
SHANK1_SA0	SHANK1_SA1

Abbreviations of brain regions

A concise summary of the major functions of the brain regions in the input dataset

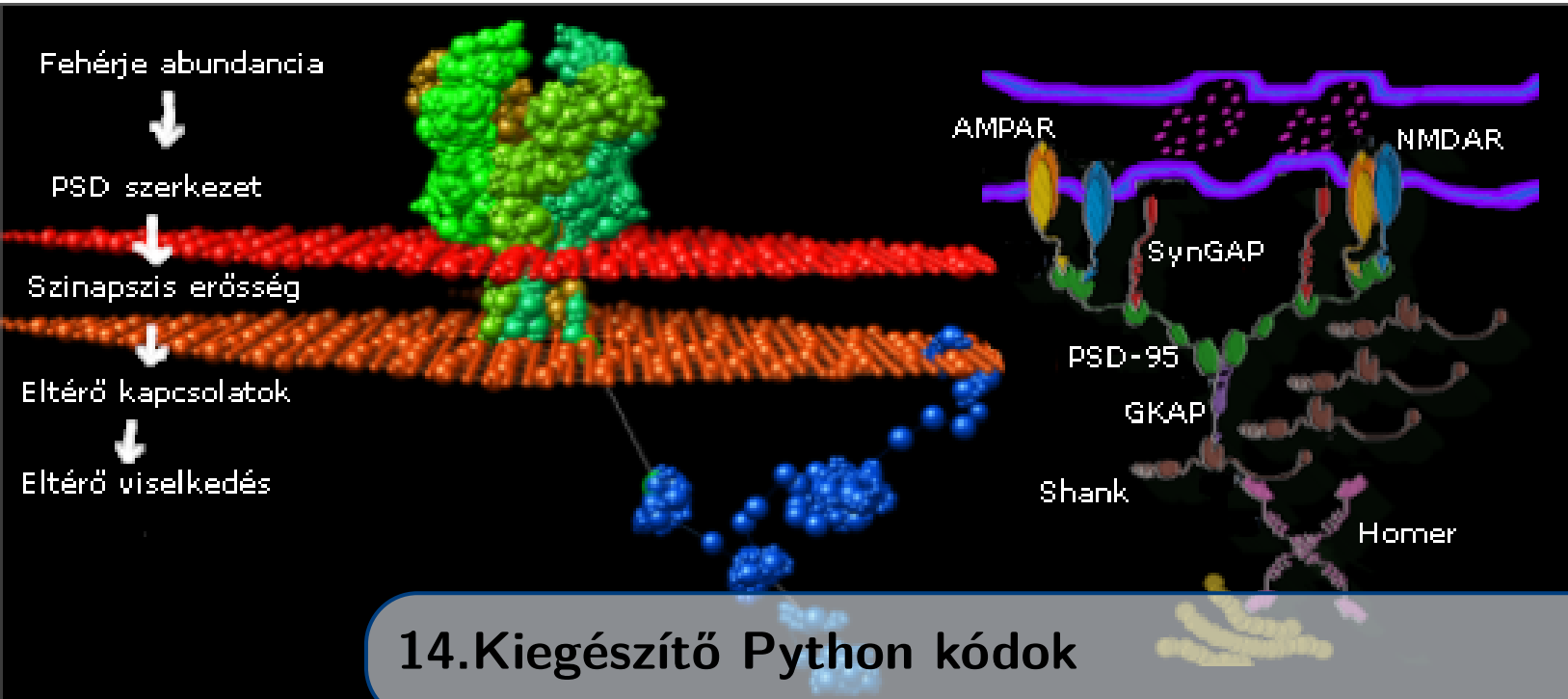
Abbreviation	{brain region}
Ocx	occipital neocortex
M1C	primary motor cortex
S1C	primary sensory cortex
AMY	amygdaloid complex
MGE	medial ganglionic eminence
STC	posterior (caudal) superior temporal cortex (area 22c)
URL	upper (rostral) rhombic lip
CGE	caudal ganglionic eminence
DTH	dorsal thalamus
MFC	anterior (rostral) cingulate (medial prefrontal) cortex
DFC	dorsolateral prefrontal cortex
OFC	orbital frontal cortex
LGE	lateral ganglionic eminence
ITC	inferolateral temporal cortex (area TEv, area 20)
HIP	hippocampus (hippocampal formation)
VFC	ventrolateral prefrontal cortex
PCx	parietal neocortex
TCx	temporal neocortex
A1C	primary auditory cortex (core)
V1C	primary visual cortex (striate cortex, area V1/17)
STR	striatum
IPC	posteroventral (inferior) parietal cortex
CBC	cerebellar cortex

Function of brain regions

Abbreviation	region function
Ocx	visual processing
M1C	pyramidal cells, motoric funtions
S1C	sensory funtions
AMY	emotions and fear
MGE	GABAergic cell production
STC	language, mediates spatial awareness, exploration
URL	granule cell production
CGE	progenitor, interneurons, striatal spiny neurons
DTH	supports adaptive decision-making
MFC	decision making, morality
DFC	social cognition (working memory, planning)
OFC	decision making
LGE	tangential mingration og GE origined cells
ITC	recognition of familiarities, spelling
HIP	memory
VFC	motor inhibition, updating plan
PCx	sensory information integraion, declarative memory
TCx	sensory input processing
A1C	sound processing
V1C	perception of visual object ,working memory
STR	coding of social action and reward
IPC	maintaining attention
CBC	

equilibrium dissociation constants found in literature Zhang et al. reconsituted PSD values

Domain 1	domain2	kd (uM)	kon	koff	reference	mutant koff	calculated
HOMER1_EVH1	SHANK1_virtualEVH1	3,00	1	3,00	23		
HOMER1_HK	HOMER1_HK		1	1,00	23		
GKAP1_CGKAP	SHANK1_PDZ	0,02	1	0,02	23	0,11	
DLG4_PDZ	GRIN2B_CNMDAR	10,00	1	10,00	23		
DLG4_PDZ	GRIA1_CAMPA	1,00	1	1,00	23		
DLG4_PDZ	SYNGAP1_CSYNGAP	0,20	1	0,20	23		
DLG4_GK	GKAP1_GBRS	0,10	1	0,10	23		
SHANK1_SA0	SHANK1_SA1		1	1,00	23		



14.Kiegészítő Python kódok

Multimerizálódott fehérjeláncok kinyerése szuperkomplexből

```
def getChains(complex,protein):
    protIndices=find_indices(complex["structure"], lambda e: e==protein)
    pairs=[]
    for binding in complex["bindings"]:
        if (binding["first_p"] in protIndices and binding["second_p"] in
            ↪ protIndices):
            pairs.append([binding["first_p"],binding["second_p"]])
    pairs=np.array(pairs)
    prots={}
    for pair in pairs:
        for prot in pair:
            if (prot in prots):
                prots[prot]+=1
            else:
                prots[prot]=1
    terminals={}
    for key in prots:
        if (prots[key]==1):
            terminals[key]=True
    chains=[]
    for key in terminals:
        pid=key
        chain=[pid]
        if(terminals[key]):
            neighbours=2
            while neighbours !=1:
```

```

    for pair in pairs:
        if (pid == pair[0] and pair[1] not in chain):
            pid=pair[1]
            chain.append(pid)
            neighbours=prots[pid]
            if (neighbours==1):
                terminals[pid]=False
        if (pid == pair[1] and pair[0] not in chain):
            pid=pair[0]
            chain.append(pid)
            neighbours=prots[pid]
            if (neighbours==1):
                terminals[pid]=False
    chains.append(np.array(chain))
chains=np.array(chains)
return chains

```

Levelek kiszedése

```

def getLeaves(complex):
    psd95s = find_indices(complex["structure"], lambda e: e == "DLG4")
    psd95sNeighbours = []
    for psd95 in psd95s:
        leaves = []
        psd95N = {}
        for binding in complex["bindings"]:
            if binding["first_p"] == psd95:
                leaves.append(complex["structure"][binding["second_p"]])
            if binding["second_p"] == psd95:
                leaves.append(complex["structure"][binding["first_p"]])
        psd95N["leaves"] = leaves
        psd95N["ID"] = psd95
        psd95sNeighbours.append(psd95N)
    return psd95sNeighbours

```

Szerkezetek összehasonlítása

```

def complexEqual2(_complex, complex_):
    if complexEqualNeighbors(_complex, complex_):
        if assignLeaves(_complex, complex_):
            if getLens(getChains(_complex, "SHANK1")) ==
            ↪ getLens(getChains(complex_, "SHANK1")):
                if assignPositions(_complex, complex_, "SHANK1", "DLGAP1"):
                    if assignPositions(_complex, complex_, "SHANK1",
                    ↪ "HOMER1"):
                        return True

```

```

        else:
            return False
    else:
        return False
    else:
        return False
    else:
        return False
else:
    return False

```

Komplex relevancia kiszámítása

```

pca2 = PCA(n_components=524)
pca2.fit(X)
eigenvectors = pca2.components_
print(len(eigenvectors))
feature_importance=np.zeros(len(eigenvectors[0]))
for i in range(len(eigenvectors)):
    feature_importance+=pca2.explained_variance_ratio_[i] * 1 /
    ↪ sum(abs(eigenvectors[i])) * abs(eigenvectors[i])
print("Feature Importance")
print(feature_importance)

with open(asksaveasfile(initialfile='feature_importance.npy',
                        defaultextension=".npy").name, 'wb') as f:
    np.save(f, feature_importance)

print(sum(feature_importance))

```

MultiPáros T-próba kiszámítása

```

def MultiPairwiseTtest():
    from google.protobuf.json_format import MessageToDict
    import scipy.stats as stats
    exp = Experiment(fd.askopenfilenames(title="Experiment Binaries"))
    print([field.name for field in exp.complexes.DESRIPTOR.fields])
    print([field.name for field in
    ↪ exp.complexes.complexes.__getitem__(0).DESCRIPTOR.fields])
    IDs = []
    i: int
    for i in range(len(exp.complexes.complexes)):
        IDs.append(exp.complexes.complexes.__getitem__(i).ID)

    with open("colorsfilename") as f:
        colorsdata = json.load(f)

```



```

colorsandtypes = colorsdata["colors"]
colors = colorsandtypes[0]
region_types_set = colorsandtypes[1]
region_names = colorsandtypes[2]
region_types = colorsandtypes[3]

avgs=[]
feature_importance = np.load(fd.askopenfilename(title="Feature
↪ Importance"))
count=0
for region in region_names:

    Ttest_of_one_condition = {}
    count+=1
    print(count)
    for i in range(len(IDs)):
        dict = {}
        for abd in
            ↪ exp.complexes.complexes.__getitem__(i).occurrences[0].abundance:
            if (abd.condition ==
                ↪ region.split("\\") [3] [:-4].replace('_',
                                                                    ↪ '-').lower()):
                    dict["WT"] = abd.values
                    break
            if "WT" not in dict:
                dict["WT"] = np.zeros(40)
            for abd in
                ↪ exp.complexes.complexes.__getitem__(i).occurrences[0].abundance:
                if (abd.condition == 'm-hom-' +
                    ↪ region.split("\\") [3] [:-4].replace('_',
                                                                    ↪ '-')):
                        dict["MT"] = abd.values
                        break
            if "MT" not in dict:
                dict["MT"] = np.zeros(40)
            test=stats.ttest_rel(dict["WT"], dict["MT"])
            if(not (np.isnan(test[0]) or np.isnan(test[1]))):
                Ttest_of_one_condition[i]=test

with open('savefilename', 'w') as outfile:
    json.dump(Ttest_of_one_condition, outfile)

sum=0
for key in Ttest_of_one_condition:

```

```
sum+=feature_importance[key]*Ttest_of_one_condition[key][1]  
avg=sum  
avgs.append(avg)
```