



Pázmány Péter Katolikus Egyetem

Roska Tamás Műszaki és Természettudományi Doktori Iskola

**Posztzinaptikus fehérjekomplexek kialakulásában szerepet
játszó interakciók bioinformatikai elemzése**

Tézisfüzet

Dobson-Kálmán Zsófia Etelka

Témavezető: Dr. Gáspári Zoltán

2022

1 Bevezetés

1.1 A posztszinapszis

Az agy a legkomplexebb szervünk, összetett funkciójáért az egymással speciális kapcsolatot kialakítani képes idegsejtek felelősek. Az agyunkat felépítő kommunikációs egységek a szinapszisok, kialakításában részt vesznek a preszinaptikus és posztszinaptikus neuronok, előbbieket az információ leadásáért, utóbbiak az információ fogadásáért felelősek. A posztszinaptikus oldal egy több ezer fehérjéből álló hálózat, ahol a fehérjék elenyésző százaléka található monomer formában, a nagy részük egymással kölcsönhat, vagy nagyobb (úgynevezett szupramolekuláris) komplexbe szerveződik - ezekből eddig több mint 220-at azonosítottak [1]. A komplexek felépítéséről egyre több kísérletes információ áll rendelkezésünkre, azonban ismereteink legnagyobb része az itt található fehérjék bináris kapcsolataiból áll, a magasabb rendű komplexekről és különösen a teljes háromdimenziós hálózat szerveződéséről igen keveset tudunk. A posztszinaptikus neuronon belül jól elkülöníthető a posztszinaptikus denzitás (PSD), amely elektronmikroszkóppal is megfigyelhető, nagy sűrűségű, félig membránhoz kötött jellegzetes struktúra. A PSD-ben legalább 1 000 különböző fehérje található meg [2]. Egy PSD átlagos tömege 1gDa, ami körülbelül 10 000 fehérje molekulának feleltethető meg [3]. Az egyes fehérjék koncentrációja is jelentős eltérést mutat, a leggyakrabban előforduló fehérje például PSD-95 (~ 300 kópia/szinapszis).

1.2 A coiled-coil motívum

A coiled-coilok az elsők között felfedezett szupermásodlagos szerkezeti elemek, ahol két vagy több alfa-hélix alkot spirált egymással úgy, hogy maximalizálják a hidrofób kontaktusok kialakulását a két lánc között [4]. Különböző organizmusokban gyakran fordulnak elő, általában a proteom 3-10%-át teszik ki élővilág szerte [5, 6]. A coiled-coil helikális szerkezetének kialakulásához egy speciális hét aminosav hosszú mintázatra van szükség, amely egymás után többször megismétlődik: ez a HPPHPPP, ahol a H hidrofób, P pedig poláris aminosavakat jelent [4]. Az egyes pozíciókat a heptádban meghatározott karakterekkel jelölik [7, 8], ahol az 'a' és a 'd' pozícióban lévő aminosavak általában hidrofóbok, míg az 'e' és 'g' pozíció a polárisak, míg a maradék három aminosav hozzáférhető a víz számára - így kiadva a heptád hét pozícióját: 'abcdefg'. Azt feltételezik, hogy szerkezet felvételében a fentieken kívül kulcsszerepe lehet egy trigger szekvenciának nevezett szakasznak is. Ez a rövid szekvencia részlet kódolhat olyan önálló feltekeredésre képes egységet, ami leszűkíti a feltekeredés terét és coiled-coil szerkezet felvételéhez segíti a fehérje következő részét [5]. A coiled-coilok között a leggyakoribb a két vagy három hélixből kialakuló szerkezetek, de magasabb szintű összerendeződés is lehetséges [9]. Attól függően, hogy hány hélix kapcsolódik a coiled-coilban, megkülönböztethetünk különböző oligomerizációs állapotokat. A láncok minden oligomerizációs állapot esetében parallel és antiparallel orientációban is előfordulhatnak [7]. A láncok származhatnak azonos fehérjékből, ilyen esetben homo-

oligomerekről beszélhetünk, vagy különböző fehérje láncokból, amiket heterooligomereknek nevezünk [6]. Bizonyos esetekben a coiled-coil egy fehérjén belüli két hélixből is összeállhat. A coiled-coilok szerkezeti diverzitása azt eredményezi, hogy változatos funkciók létrehozását teszik lehetővé [10]. Gyakran előfordulnak motor fehérjékben, transzkripciós faktorokban és receptorokban.

1.3 Fehérje-fehérje interakciók

A fehérje interakciók legalább két fehérje fizikai kontaktusba lépését jelentik molekuláris dokkolás révén, a létrejövő kapcsolat sejtekben vagy élő organizmusokban *in vivo* fordul elő [11]. Az interakció révén kialakuló fizikai kölcsönhatás specifikus, valamilyen funkció ellátására evolválódik [11]. A sejtes környezetben számos egymással versengő interakció áll fenn, még ha ezek közül a legtöbb kötés csak nagyon gyenge kapcsolatot tud létrehozni a fehérjék között. Azonban a kötés tényleges létrejöttének valószínűségét nagyobb mértékben befolyásolja az adott fehérjék térbeli és időbeli eloszlása (azaz, hogy egy időben és egy helyen jelenjenek meg), mint a kötés erőssége. Az interakciók kialakításában többféle fehérje szakasz is részt vehet: domének és rövidebb-hosszabb rendezetlen szakaszok. A különböző interakciós lehetőségek közül a legtöbbet a domének révén kialakított kötésekről tudunk. Az interakciós domének tipikusan 35-150 aminosav hosszúságúak. A jelenleg ismert domének összesen több százezer bináris interakciót és $\sim 10\,000$ komplexet képesek létrehozni [12], [13]. A doménekkal szemben a rendezetlen régiók funkcionális előnye, hogy a szerkezeti flexibilitás miatt képesek több, extrém esetekben akár több száz partnerhez is kötődni [14]. A rövidebb rendezetlen szakaszok, lineáris motívumok is számos interakció kialakításáért felelősek, azonban ezen kapcsolatok általánosságban jóval kevésbé feltártak. Jelenlegi elképzelésünk szerint akár egy millió ilyen interakció is előfordulhat eukarióta sejtekben, de ezek töredékét igazolták csak kísérletesen [12]. A fehérje-fehérje interakciók eltérő affinitással alakulnak ki. A spektrum egyik végén a rövid életű, alacsony affinitású, tranziens interakciók vannak (ezek leggyakrabban rövid lineáris motívumok), amelyeknél a kötés milliszekundumnyi idő alatt lejátszódik majd megszűnik, és a kötés erőssége mikro- vagy millimoláris tartományba esik [15]. A spektrum másik oldalán a szorosabb kölcsönhatások találhatóak (tipikusan domén-domén), amelyek felezési ideje a több órás tartományba is eshet, és akár nano- vagy pikomoláris affinitással is megvalósulhatnak. A fehérje-fehérje interakciók kimutatására számos lehetőség van. A különböző technikák eltérő adatokat szolgáltatnak az interakciókról, emellett a megbízhatóságuk is eltérő egy-egy hipotézis megerősítésére.

1.4 Betegséget okozó csírvonal mutációk

A replikáció az egyik legfontosabb funkció a sejtekben. Ez a folyamat elképesztően precíz, a mutációk gyakorisága (azaz a hibák előfordulása) 10^{-10} mutáció/bázispár sejtosztódásonként [16]. Abban az esetben, ha ezek a hibák csírasejt-vonalban fordulnak elő, akkor az utódoknak és azok minden sejtjének is átadódnak [17, 18]. Összességében a mutációk hatása, fenotípusos megnyilvánulása egy spektrumon képzelhető el. Minden genetikai variáció mutációkból indul ki, a mutációk egy része negatív következményeket hordoz magával a sejtekre nézve, míg mások előnyt fognak biztosítani és a szelekciót pozitívan befolyásolják az evolúciós fejlődés során [17]. Lesznek tehát kifejezetten előnyös mutációk, amik a spektrum pozitív végén helyezkednek el, ezek szelekciós előnnyel fognak járni. Léteznek olyan mutációk, amelyek bár fenotípusos változást okoznak, semleges hatásúak lesznek. Ezeket a köznyelvben is általában polimorfizmusnak nevezzük (pl. szemszín). A spektrumon tovább haladva következnek a betegséget okozó mutációk, amelyek fenotípusos megnyilvánulása negatívan érinti a sejtet. Ezek a mutációk eltérő hatással lesznek a sejtekre (pl. laktóz intoleranciától kezdve a Huntington betegségig és tovább). A csírvonalbeli betegséget okozó mutációk általában egy gyengébb fenotípust eredményeznek, ami negatívan befolyásolja az életminőséget, de nem végzetes.

2 Módszerek

A doktori munkám során szabad felhasználású módszereket és adatokat használtam. Az adatok feldolgozására Python scripteket írtam elsősorban Linux operációs rendszer alatt. A disszertációban bemutatott lépések során több forrást és technikát is igénybe vettem, a legfontosabbak a következők voltak:

Adatbázisok:

Uniprot: <https://www.uniprot.org/>

PDB: <https://www.rcsb.org/>

Pfa.: <https://pfam.xfam.org/>

SynaptomeDB: <http://metamoodics.org/SynaptomeDB/index.php>

G2C: <https://www.genes2cognition.org/>

SynGO: <https://www.syngoportal.org/>

BioGRID: <https://thebiogrid.org/>

Intact: <https://www.ebi.ac.uk/intact/home>

STRING: <https://string-db.org/>

Módszerek:

BLAST: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

CD-HIT: <https://www.bioinformatics.org/cd-hit/>

ClustalO: <https://www.ebi.ac.uk/Tools/msa/clustalo/>

GO: <http://geneontology.org/>

DeepCoil: <https://toolkit.tuebingen.mpg.de/tools/deepcoil>

Ncoils: <https://predictprotein.org/>

Paircoil: <http://cb.csail.mit.edu/cb/paircoil/>

Marcoil: <https://toolkit.tuebingen.mpg.de/tools/marcoil>

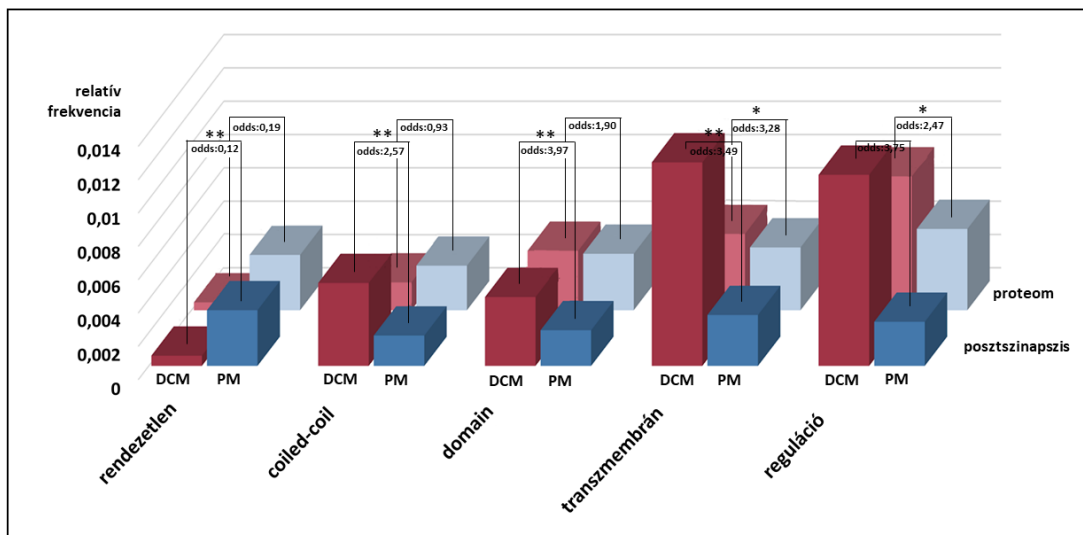
Logicoil: <http://coiledcoils.chm.bris.ac.uk/LOGICOIL/>

3 Új tudományos eredmények

3.1 Első tézis csoport: coiled-coil szerkezeti elemek vizsgálata

3.1.1 Megállapítottam, hogy a betegséget okozó csírvonal mutációk a semleges polimorfizmusokhoz viszonyítva feldúsulnak coiled-coil régiókban a posztzsinapszis fehérjéi esetében, míg a teljes humán proteomban nem. A rendezetlen és a transzmembrán régiók esetén a frekvenciák eloszlása azonos a posztzsinapszis és a proteom fehérjéiben.

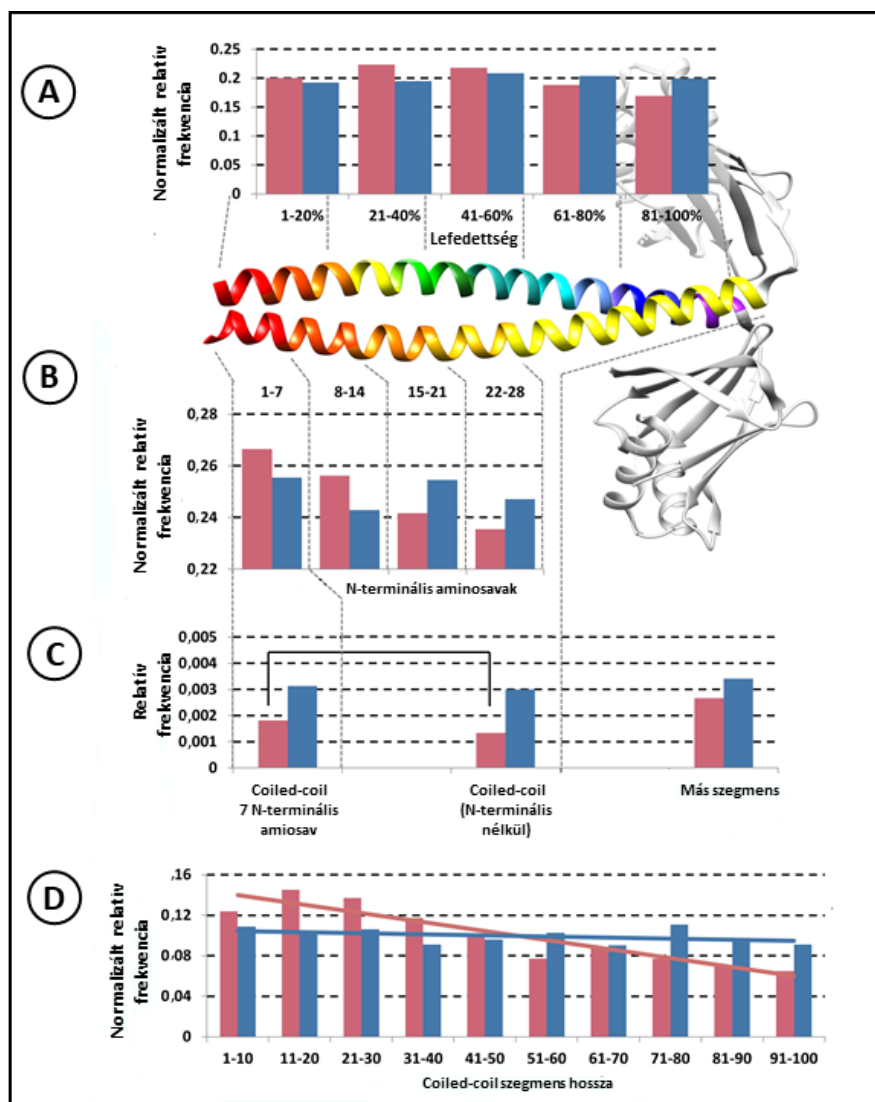
A posztzsinaptikus (PS) fehérjék szerkezeti kitétségének vizsgálatához meghatároztam a mutációk eloszlását a predikciókkal meghatározott szerkezeti és funkcionális régiókban, mind a posztzsinaptikus fehérjék ('PS_STRICT' adatszett) mind a teljes humán proteom esetében. Vizsgáltam a transzmembrán (CCTOP), coiled-coil (Deepcoil, Ncoils, Marcoil és Paircoil), rendezetlen (IUPred predikció alapján) és a domén (Pfam) részeket (ellentmondás esetén ebben a prioritási sorrendben), illetve a fehérjék szabályozásában részt vevő, poszttranszlációs módosításoknak kitett aminosavakat (PhosphoSitePlus). Megfigyelhető volt, hogy a rendezetlen szakaszokon a polimorfizmusok (PM) feldúsulnak a betegséget okozó mutációkkal (DM) szemben mind a PS, mind a teljes proteom fehérjéi esetében (1. ábra). Ezzel szemben a DM-ek frekvenciája általában véve magasabb a többi, nem rendezetlen szerkezeti részben. Ezek a tendenciák mind a két adatszettben jellemzőek, azonban a PS esetében mindkét hatás markánsabban jelenik meg. Egyetlen esetet találtunk, ahol a teljes proteomban és a posztzsinapszisban észlelhető trendek ellentétesek: a coiled-coil régiókban a PS esetében magasabb a DM-ek relatív gyakorisága, mint a polimorfizmusoké (1. ábra). Két szempontból is érdekesnek találtuk ezt az eredményt: egyrészt csak ebben az esetben láttunk teljesen ellentétes tendenciát a posztzsinapszis és a proteom között, másrészt a coiled-coil eredményeknek szerkezeti szempontból kiindulva sokkal inkább a többi rendezett részen (globuláris domén vagy a transzmembrán régiókban) történő dúsuláshoz kellene hasonlítania, mint a rendezetlen adatokhoz. Ennek oka, hogy a DM-ek sokkal nagyobb kárt tudnak okozni egy kompakt szerkezetben annak megbontásával, mint a nagy konformációs szabadsággal rendelkező flexibilis szakaszokon, ahol "nincs szerkezet" ami megváltozhatna, 'elromolhatna'. A jelenség mélyebb megértésének érdekében tovább vizsgáltuk a DM-ek szerepét a coiled-coil régiókban.



1. ábra A betegséget okozó mutációk (DM) és polimorfizmusok (PM) relatív gyakorisága a PS és a proteom fehérjéiben, az egyes szerkezeti és funkcionális csoportokban (x tengely). A posztiszinapsztikus adatszetten számolt frekvenciák esetében a DM vörössel, PM sötétkékkel színezett, míg a proteom esetében előbbiek halvány pirosak, utóbbiak halványkékek

3.1.2 Kimutattam, hogy a coiled-coil N-terminális régiókban magasabb a betegséget okozó mutációk relatív gyakorisága, mint a coiled-coil régiók többi részén, ami ezen régió esszenciális szerepét sejteti.

Bár a DM-ek frekvenciája a coiled-coil régiókon kívül magasabb, így is számos mutáció esik közvetlenül coiled-coil régióba. Első lépésként a coiled-coil régiókat hosszuk mentén adott számú szakaszra osztottuk százalékos arányokat használva (1-20%, 21-40%, 41-60% stb.) (2.A ábra). Ezen megközelítés szerint nem látszódott egyértelműen eltérő dúsulás a különböző régiók között (legfelső panel). Ezzel szemben, ha a coiled-coil régiókat a kanonikus heptád ismétlés mentén osztottuk fel (1-7, 8-14, stb.) (2.B ábra), egyértelműen megfigyelhető volt, hogy a DM-ek az N-terminális 7 aminosavat érintik a legerősebben. A megközelítés előnye, hogy a coiled-coil régiók hosszbeli eltéréseiből származó aránytalanságok is kiküszöbölhetővé váltak. Az N-terminális 7 aminosav a coiled-coil többi részével összehasonlítva is szignifikáns halmozódást mutatott (χ^2 teszt, $p < 0.01$) és az arányhányados (odds ratio) 1,33 volt a DM-ek és a PM-ek között (2.C ábra). Ez a hatás annyira erős, hogy a rövidebb coiled-coilokban is magasabb a DM-ek relatív gyakorisága, míg a polimorfizmusok hosszától függetlenül teljesen egyenletes eloszlást mutatnak (2.D ábra). Ennek egy lehetséges oka lehet, hogy az N-terminális régióknak kiemelt szerepe lehet a feltekeredésben és akár a trigger szekvencia is lokalizálódhat itt.



2. ábra A DM-ek és PM-ek eloszlása a coiled-coil szerkezetekben és szerkezeti részeken kívül A: A DM-ek és PM-ek normalizált frekvenciákra a coiled-coil régió hosszában - százalékos felosztás alapján B: A DM-ek és PM-ek eloszlása az regiszterek alapján meghatározott csoportokban C: A DM-ek halmozódása az N-terminális 7 aminosavban a coiled-coil régió többi részével összehasonlítva D: A DM-ek halmozódása a coiled-coil szegmensek hosszának függvényében (DM:piros, PM:kék)

3.1.3 A DiseaseOntology adatait felhasználva megállapítottam, hogy a betegséget okozó mutációk jellemzőek azon coiled-coil tartalmú fehérjékben, amelyek összefüggésbe hozhatóak idegrendszeri megbetegedésekkel.

A coiled-coil szerkezetek és az ezeket tartalmazó fehérjék rendkívül különböző funkciókat töltenek be a sejtekben. Az, hogy az ide eső mutációk milyen betegségeket okoznak, megmutathatja, hogy milyen funkciók ellátásához kritikus a coiled-coil szerkezeti részek integritása. A Disease Ontology (DO) ontológia rendszer a humán betegségek leírására készült, ahol az egyes betegségek specifikusan cso-

portosítva vannak egy hierarchikus rendszerben. A DO, valamint a Humsavar adatait összepárosítva meghatározható, hogy egy pontmutáció milyen betegség osztályhoz/osztályokhoz köthető. Ez azért fontos, hiszen így nem csak egyedi fehérje szinten vizsgálhatjuk a fehérjéket, illetve nem is önkényesen definiált csoportokban, hanem releváns, szakértők által összeállított rendszerekben értelmezhetőek az adatok.

Egy betegség azonban több útvonalon is megtalálható, például a Alzheimer-kór genetikai betegségek közé is sorolt, illetve anatómia entitás szerint is hozzárendelt. Az elemzésből az látszik, hogy a legtöbb általunk megtalált, mutációkkal érintett coiled-coil fehérje központi idegrendszeri betegséggel volt összefüggésbe hozható, emellett jelentősen megtalálhatóak izom, szenzoros és bőr betegségek esetében is betegséget okozó mutációk által érintett fehérjék (3. ábra).

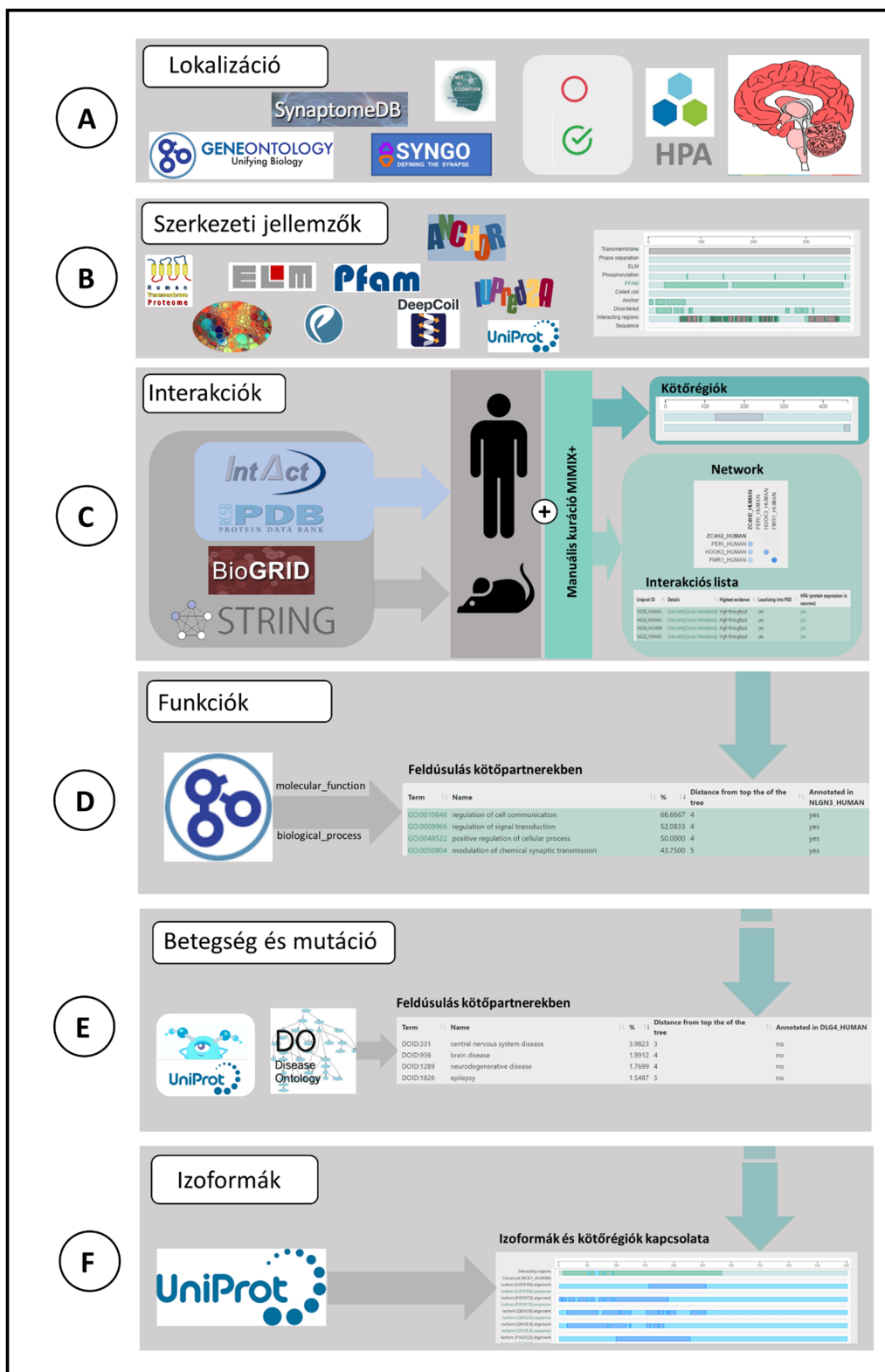
DO szint1	Fehérje szám	DO szint2	Fehérje szám	DO szint3	Fehérje szám
betegség anatómiai entitás alapján	734	kültakarót érintő betegség	72	bőr betegség	71
		mozgásszervi betegség	165	izom betegség	161
		idegrendszeri betegség	397	központi idegrendszeri betegség	294
				szenzorosbetegség	74
kardiovaszkuláris betegség	60	szív betegség	60		
szindróma	52				
genetikus betegség	56				
metabolikus betegség	103	örökölt metabolikus betegség	63	karbohidrát metabolizmus bet.	63

3. ábra Betegség osztályok a megfelelő DiseaseOntology szint jelölésekkel és a mutációk által érintett fehérjék száma

3.2 Második tézis csoport: A posztszinapszis interakcióinak vizsgálata

3.2.1 Létrehoztuk a PostSynapticInteractionDataBase-t, amelyben az általam gyűjtött adatokat kiegészítettük más interakciós, szerkezeti és funkciós adatokkal és az adatbázist hozzáférhetővé tettük.

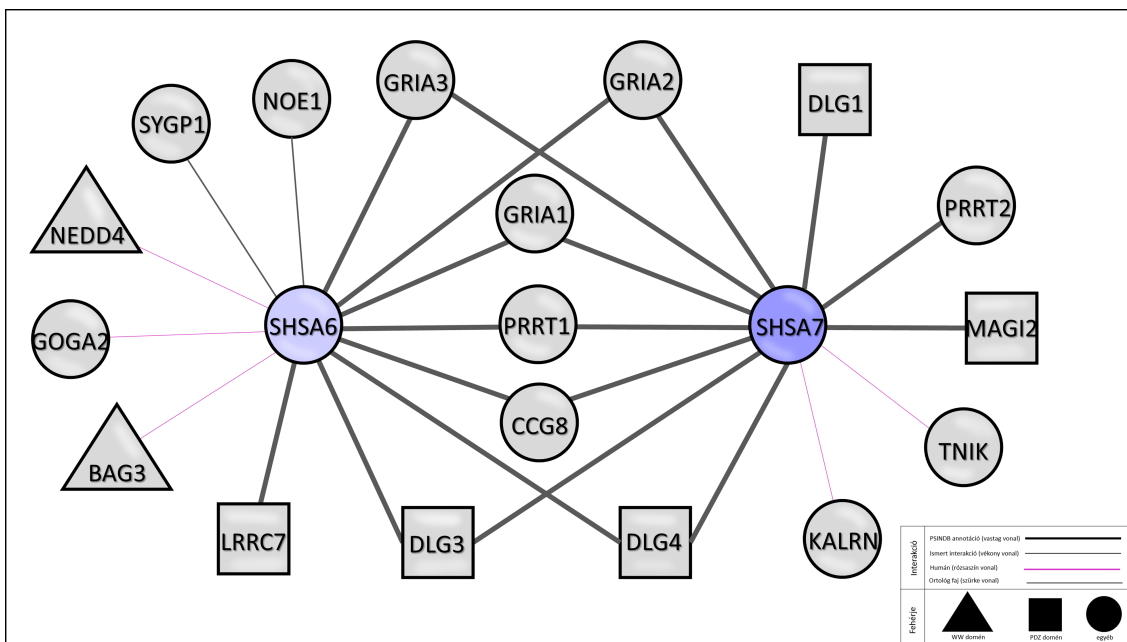
Az alapvető célunk az adatbázis felállításánál egy olyan posztszinapszis specifikus bináris fehérje-fehérje kölcsönhatásokat (protein-protein interaction, PPI) tartalmazó gyűjtemény létrehozása volt, amely az PPI-k mellett az interakciókat befolyásoló és meghatározó szerkezeti és funkcionális adatokat is részletesen tartalmazza. Ehhez manuális adatfeldolgozást használva nagy mennyiségű interakciós adatot gyűjtöttem egy előre meghatározott annotációs rendszerben. Az interakciós adatok esetében a manuális kiértékelés mellett négy adatbázis adatainak integrálása történt meg: IntAct, BioGRID, STRING, valamint a PDB-ben található több alegységű fehérjék kölcsönhatásai. A fehérjék interakciós lehetőségeit alapvetően meghatározzák a bennük megtalálható szerkezeti és funkcionális elemek. Ezek meghatározásához régióspecifikus predikciókat és adatokat gyűjtöttünk össze, amelyeket a szekvencia mentén jelölünk: transzmembrán topológia (HTP adatbázisból átvéve), coiled-coil (DeepCoil predikció és UniProt annotáció), rendezetlen régiók / rendezetlen kötő régiók (IUPred2A), fázis szeparáció (PhasePro), domének (Pfam). Az interakciós és szerkezeti adatok mellett funkcionális információkkal egészítettük ki az adatbázist, ezek közé tartoztak a betegséget okozó mutációk és polimorfizmusok (Humsavar), a molekuláris funkció és biológiai folyamat terminusok (GO) és betegség terminusok (DiseaseOntology). Emellett a rövid lineáris motívumokat (Eukaryotic Linear Motif adatbázis) és az fehérje izoformáit (UniProt) valamint foszforilációs adatokat (szintén UniProt) is integráltuk az adatbázisban. Az adatbázis fizikailag a PPKE ITK szerverein található, elérhetősége: <https://psindb.itk.ppke.hu> (4. ábra).



4. ábra A PSINDB adatrétegei A: A poszt-szinaptikus lokalizáció (forrás: SynptomeDB, G2C, SynGO, GO) és expresszió (forrás HumanProteinAtlas), B: A fehérjék szerkezeti és funkcionális jellemzői (főbb források: IUPRED2A, Pfam, UniProt, részletekért lásd a szöveget), C: Interakciók, D: Funkciók, E: Betegség és mutáció, F: A fehérje izoformák és kötőrégiók kapcsolata

3.2.2 Kidolgoztam egy kurációs rendszert, amely a posztszinaptikus fehérje-fehérje interakciók leírásának legjobb gyakorlatait követi és aminek segítségével több, mint kétezer kísérletes adatot rendszereztem és amelynek segítségével eddig ismeretlen alhálózatot tártam fel.

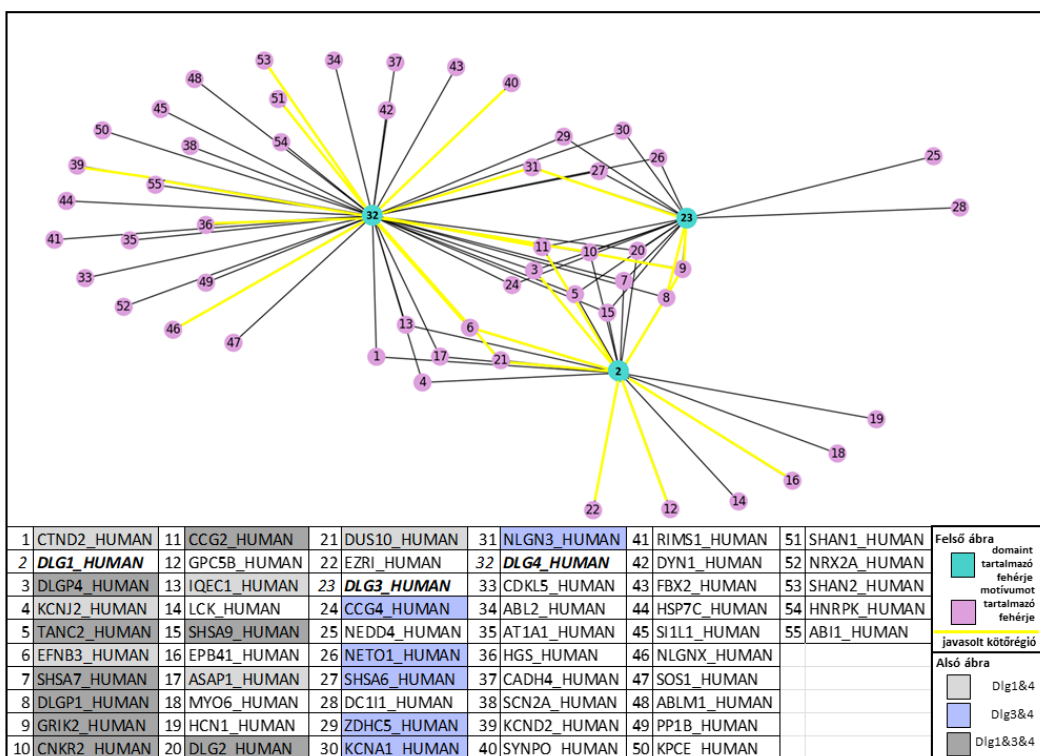
A PSINDB felállításában velem együtt több olyan kutató is részt vett, akinek volt már tapasztalata az IntAct rendszerével és együttesen arra a következtetésre jutottunk, hogy a számunkra releváns információk tárolására első megközelítés szerint egy kiterjesztett MIMIx formátum is megfelelő lesz (MIMIx+). Ebben a rendszerben a MIMIx-ben meghatározott szükséges információk mellett az interakcióban (feltételezhetően) résztvevő kötőrégiók is rögzítésre kerülnek. A HUPO-PSI ajánlásokat más területen is követtük. A kísérletek és azzal összefüggő információk leírásához a PSI-MI kontrollált szókészletet (Controlled Vocabulary) és a kereséshez az EBI ontológiakereső oldalát, az Ontology Lookup Service-t (OLS) használtuk. Végül, de nem utolsósorban a PSI által ajánlott PSI-MI formátumban tettük közzé az adatainkat. A PSINDB használható lokális alhálózatok részletes vizsgálatára: a SHSA6 és SHSA7 fehérjék bitopikus transzmembrán fehérjék. Az AMPA receptorok számos "támogató" fehérjével állnak kapcsolatban, amelyek intracelluláris szállításban és az AMPA funkcionális sajátosságainak kialakításában vesznek részt. Ezen fehérjék közé tartoznak az SHSA fehérjék is, A fehérje család első tagját, az SHSA9 fehérjét 2010-ben fedezték csak fel [19]. A Uniprot funkcionális leírása szerint a hippokampusz C1, C3 régióinak szinapszisaiban töltnek be szerepet a transzmisszió fenntartásában, tehát meghatározó szereplői a szinapszisnak. A SynGO osztályozása alapján a posztszinaptikus receptor aktivitás regulációjában van szerepük. A két (homológ) fehérje esetében már korábban is feltártak interakciós partnereket (6 partner a SHSA6, 2 partner a SHSA7 esetében), azonban a PSINDB-hez hozzáadott manuális kuráció eredményeként, előbbi esetében 8, utóbbi esetén 11 partnerrel sikerült kiegészítenem a hálózatot (5. ábra). Mind az SHSA6, mind a SHSA7 fehérjében nagy valószínűség szerint olyan motívumok találhatóak, amelyek segítségével az általam azonosított partner fehérjékhez kapcsolódni képesek. Az extrém C-terminális régióban egy 2-es típusú PDZ kötő motívum található, amelyet SHSA6 esetében kísérletesen is igazoltak már [20]. Az újonnan feltárt partnereik között öt fehérjében található meg PDZ domén. Emellett nagy valószínűséggel tartalmaznak WW domén kötő motívumot is, amit az is alátámaszt, hogy a SHSA6 esetében két kötő partnerben is megtalálható ilyen domén. Az SHSA6 esszenciális szerepét a procedurális memória kialakulásban nemrég írták le [21], illetve kimutatták a fehérjét kódoló gén érintettségét kognitív zavarban [22]. A PSINDB adataira támaszkodó elemzés eredményeként egy jóval komplexebb képet kaptunk ezen fehérjék neuronális szerepéről. Ez a példa jól mutatja, hogy az adatok integrálása egy adatbázisban segítheti a rendszerek bővebb megértését. Emellett rámutathat a rendszer azon pontjaira, ahol a jelenlegi adatok tovább bővíthetőek a kísérletes adatok gyűjtésével és rendszerezésével.



5. ábra SHSA6 és SHSA7 fehérjék PSINDB által kibővített hálózata interakciós hálózata Az interakciókat a vonalak reprezentálják; jelölés: vékony: már ismert, vastag: új annotáció, rózsaszín: humán taxon, szürke: ortológ A fehérjék jelölése: négyzet: PDZ domént tartalmaz, háromszög: WW domént tartalmaz, kör: egyéb domént tartalmaz

3.2.3 A PSINDB adatait felhasználva javaslatot tettem domén-domén és domén-motívum interakciókat létrehozó kötőrégiókra.

A PSINDB-ben található interakciók nagy hányadában nem ismert a két fehérje pontos fizikai kapcsolódásának a helye. Ennek oka lehet, a forrás adatbázis adat gyűjtési döntése (BioGRID), illetve, ha az interakció predikcióból származik (STRING). Emellett az ortológ fajokból származó kötőrégiókat mi sem vesszük át az IntActból. A posztzsinaptikus működés és a kapcsolatok feltárásának azonban fontos része a konkrét kötések megismerése. A domén-domén interakciók alapvetően egy fontos területét jelentik a szinapszis kutatásoknak, azonban a tranziens motívum-domén kölcsönhatásokról viszonylag kevés irodalmi adat áll rendelkezésünkre. Továbbá a PS-ben kialakuló komplexek lehetséges összerendezéséről képet kaphatunk például a kompetitív kötőhelyek és partnerek meghatározásával. Ezért két megközelítés szerint, domén-domén és domén-motívum interakciók kötő régióira tettem javaslatot. A Dlg1, Dlg3 és Dlg4 fehérje példájából az látszik, hogy a doméneken keresztül megvalósuló kölcsönhatások esetében a Dlg1 esetében az interakciós partnerek 11%-ra tudunk kötőrégiókra javaslatot tenni, míg a Dlg3 esetében ez 8%, a Dlg4 esetében 6%. Viszonyításként az előbb említett sorrendben a már ismert kötőrégiók 18%-át, 2%-át és 7%-át teszik ki az kötéseknak. A DLG1, DLG3 és a DLG4-ben olyan domének találhatóak, amelyek “tipikus” motívum kötő doménnek számítanak (pl. SH3 vagy PDZ). Abban az esetben, ha ezek a fehérjék a domén oldalon találhatóak és a partnerükben



6. ábra ábra A Dlg1, Dlg3, Dlg4 fehérjék domén-motívum kölcsönhatásai Felső panel: A feltárt domén-motívum hálózat (kék: domént tartalmazó fehérje, rózsaszín: motívumot tartalmazó fehérje, sárga, javasolt kötőrégiójú kapcsolattal rendelkező kapcsolatok) Alsó panel: A hálózati számozott jelölések feloldása(a három fehérje közös partnereinek jelölése, világos szürke: Dlg1 és Dlg4, kék: Dlg3 és Dlg4, illetve sötétszürke: Dlg1, Dlg3 és Dlg4)

megtalálható a lineáris motívum, 152 új kötőrégióra lehet javaslatot tenni (továbbra is alkalmazott kényszerfeltételek, rendezetlen=1, hozzáférhetőség=1, shannon=0,0). Ebből 50 esetben nem ismert humán régió és 18 esetben ismert a régió, de az máshol található mint a pipeline által javasolt szakasz (6. ábra).

4 Összefoglalás

A posztszinapszis a neuronális működés fontos egysége, amely molekuláris mechanizmusai olyan alapvető agyi funkciók elemei, mint a memória vagy a tanulás. Bár ismereteink nagy mértékben bővültek az elmúlt évtizedekben, számos az organellum megéréséhez elengedhetetlen információ még nem áll birtokunkban. A posztszinapszis működésének alapját a posztszinaptikus fehérjék adják, amelyek szinte elképzelhetetlen bonyolultságú hálózatok létrehozása révén fejtik ki biológiai funkciójukat.

Ezen bonyolult hálózat jövőbeli megértéséhez tesz hozzá reménység szerint disszertációm, amelyben *in silico* módszerekkel több aspektusból vizsgáltam a posztszinapszis rendszerét, elsősorban azonban a fehérjék bináris interakcióira, az ezekben szerepetjátszó funkcionális elemekre és mutációikra fókuszálva. Először a csírvonal mutációk hatását vizsgáltam a posztszinaptikus fehérjék szerkezeti elemein. Egy érdekes eredmény - a betegséget okozó mutációk előfordulása a fehérjékben - és annak alacsony irodalmi leírása arra sarkalt bennünket, hogy részletesebben is megvizsgáljuk a csírvonal mutációk hatását a coiled-coil szerkezetekben. Ennek analízisét azonban már kiterjesztettem a teljes humán proteomra. Érdekes módon a betegségek vizsgálatánál a központi idegrendszer érintettsége is markánsan látszódott. A disszertációm második felében egy posztszinaptikus interakciós adatbázis (PSINDB) felállítását mutatom be, amely egyedi és az interakciók mellett releváns biológiai adatokat tartalmazott. Adatbázisunk elsődleges adatbázisnak is tekinthető, mivel ~ 2000 interakciós adatot tartalmaz, amely nem szerepel meghatározó interakciós adatbázisokban. A manuális kuráció során a terület legjobb gyakorlatait alkalmaztuk. Emellett a PSINDB adatainak elemzéséből származó észrevételeket felhasználva jósultunk domén-domén, illetve domén-motívum kölcsönhatásokat. Mindkét interakciónak kiemelkedő jelentősége lehet a posztszinapszis funkciójában, és ezek egy része még biztosan felderítetlen.

A posztszinapszis működésének megértése nem csak azért lenne fontos, hogy tisztában legyünk alapvető agyi funkcióink molekuláris mechanizmusával, hanem azért is, hogy megértsük azt a több, mint százötven neuronális betegséget, amelyet az organellum fehérjével hoztak már összefüggésbe.

5 Publikációk

Tézis alapjául szolgáló

Kalman, Zs. E., Mészáros, B., Gáspári, Z., Dobson, L. (2020). Distribution of disease-causing germline mutations in coiled-coils implies an important role of their N-terminal region. *Scientific reports*, 10(1), 1-12.

Kalman, Zs. E., Dudola, D., Mészáros, B., Gáspári, Z., Dobson, L. (2022). PSINDB: the postsynaptic protein-protein interaction database. *Database*, 2022, baac007.

Egyéb

Quaglia, F., Mészáros, B., Salladini, E., Hatos, A., Pancsa, R., Chemes, L. B., ... Piovesan, D. (2022). DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Research*, 50(D1), D480-D487.

Frank, K., Bana, N. Á., Bleier, N., Sugar, L., Nagy, J., Wilhelm, J., ... Steger, V. (2020). Mining the red deer genome (*Cervus elaphus*) to develop X- and Y-chromosome-linked STR markers. *PLoS One*, 15(11), e0242506.

Nem lektorált folyóiratban megjelent

Kalman, Zs. E., Gáspári, Z. (2021). A preliminary study on the cistrome of human postsynaptic density from an evolutionary and network-based perspective. *bioRxiv*.

Irodalom jegyzék

- [1] Seth GN Grant. Synapse molecular complexity and the plasticity behaviour problem. *Brain and Neuroscience Advances*, 2:2398212818810685, 2018.
- [2] Takeshi Kaizuka and Toru Takumi. Postsynaptic density proteins and their involvement in neurodevelopmental disorders. *The Journal of Biochemistry*, 163(6):447–455, 2018.
- [3] Morgan Sheng and Eunjoon Kim. The postsynaptic organization of synapses. *Cold Spring Harbor perspectives in biology*, 3(12):a005678, 2011.
- [4] Derek N Woolfson. The design of coiled-coil structures and assemblies. *Advances in protein chemistry*, 70:79–112, 2005.
- [5] Jody M Mason and Katja M Arndt. Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem*, 5(2):170–176, 2004.
- [6] Linda Truebestein and Thomas A Leonard. Coiled-coils: The long and short of it. *Bioessays*, 38(9):903–916, 2016.
- [7] Andrei N Lupas, Jens Bassler, and Stanislaw Dunin-Horkawicz. The structure and topology of α -helical coiled coils. *Fibrous Proteins: Structures and Mechanisms*, pages 95–129, 2017.
- [8] Gevorg Grigoryan and Amy E Keating. Structural specificity in coiled-coil interactions. *Current opinion in structural biology*, 18(4):477–483, 2008.
- [9] Oliver D Testa, Efrosini Moutevelis, and Derek N Woolfson. Cc+: a relational database of coiled-coil structures. *Nucleic acids research*, 37(suppl_1):D315–D322, 2009.
- [10] Andrei N Lupas and Markus Gruber. The structure of α -helical coiled coils. *Advances in protein chemistry*, 70:37–38, 2005.
- [11] Zeeshan Shaukat, Sara Aiman, Chun-Hua Li, et al. Protein-protein interactions: Methods, databases, and applications in virus-host study. *World Journal of Virology*, 10(6):288, 2021.
- [12] Peter Tompa, Norman E Davey, Toby J Gibson, and M Madan Babu. A million peptide motifs for the molecular biologist. *Molecular cell*, 55(2):161–169, 2014.
- [13] Erica A Golemis, Erica Golemis, and Peter David Adams. *Protein-protein interactions: a molecular cloning manual*. CSHL Press, 2005.
- [14] Alexander Cumberworth, Guillaume Lamour, M Madan Babu, and Jörg Gsponer. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochemical Journal*, 454(3):361–369, 2013.

- [15] Xuan Yang and Andrey A Ivanov. Computational structural modeling to discover ppi modulators. In *Protein-Protein Interaction Regulators*, pages 87–108. 2020.
- [16] Bradley D Preston, Tina M Albertson, and Alan J Herr. Dna replication fidelity and cancer. In *Seminars in cancer biology*, volume 20, pages 281–293. Elsevier, 2010.
- [17] Catarina D Campbell and Evan E Eichler. Properties and rates of germline mutations in humans. *Trends in Genetics*, 29(10):575–584, 2013.
- [18] Yanmei Dou, Heather D Gold, Lovelace J Luquette, and Peter J Park. Detecting somatic mutations in normal cells. *Trends in Genetics*, 34(7):545–557, 2018.
- [19] Mazyar Abdollahi Nejat, Remco V Klaassen, Sabine Spijker, and August B Smit. Auxiliary subunits of the ampa receptor: The shisa family of proteins. *Current Opinion in Pharmacology*, 58:52–61, 2021.
- [20] Remco V Klaassen, Jasper Stroeder, Françoise Coussen, Anne-Sophie Hafner, Jennifer D Petersen, Cedric Renancio, Leanne JM Schmitz, Elisabeth Normand, Johannes C Lodder, Diana C Rotaru, et al. Shisa6 traps ampa receptors at postsynaptic sites and prevents their desensitization during synaptic activity. *Nature communications*, 7(1):1–12, 2016.
- [21] Saša Peter, Bastiaan HA Urbanus, Remco V Klaassen, Bin Wu, Henk-Jan Boele, Sameha Azizi, Johan A Slotman, Adriaan B Houtsmuller, Martijn Schonewille, Freek E Hoebeek, et al. Ampar auxiliary protein shisa6 facilitates purkinje cell synaptic excitability and procedural memory formation. *Cell reports*, 31(2):107515, 2020.
- [22] Jairo Ramos, Laura J Caywood, Michael B Prough, Jason E Clouse, Sharlene D Herington, Susan H Slifer, M Denise Fuzzell, Sarada L Fuzzell, Sherri D Hochstetler, Kristy L Miskimen, et al. Genetic variants in the shisa6 gene are associated with delayed cognitive impairment in two family datasets. *Alzheimer’s & Dementia*, 2022.