DOI:10.15774/PPKE.ITK.2023.001

Pázmány Péter Katolikus Egyetem

Roska Tamás Műszaki és Természettudományi Doktori Iskola

Posztszinaptikus fehérjekomplexek kialakulásában szerepet játszó interakciók bioinformatikai elemzése

PhD disszertáció

Dobson-Kálmán Zsófia Etelka

Témavezető: Dr. Gáspári Zoltán

2022

DOI:10.15774/PPKE.ITK.2023.001

Tartalomjegyzék

DOI:10.15774/PPKE.ITK.2023.001

DOI:10.15774/PPKE.ITK.2023.001

DOI:10.15774/PPKE.ITK.2023.001

DOI:10.15774/PPKE.ITK.2023.001

Rövidítésjegyzék

Rövidítés	Angol megfelelő	Magyar megfelelő
AMPAR	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor	(2-amino-3-(5-metil-3-oxo-1,2-oxazol-4-il)propánsav) glutaminsav-receptor
AS/AA	Aminoacid	aminosav
ASA	Solvent Accessible Surface Area	oldószerrel hozzáférhető felület
BLAST	Basic Local Alignment Search Tool	-
BRET	Bioluminescence Resonance Energy Transfer	biolumineszencia rezonancia energia transzfer
CAMKII	Calcium/Calmodulin Dependent Protein Kinase II	kálcium/kalmodulin-függő proteinkináz-II
CC	Coiled-Coil	coiled-coil
CD	Circular Dichroism	cirkuláris diklorizmus
CNV	Copy Number Variation	kópiaszám variációk
co-IP	Co-Immunoprecipitation	Ko-immunoprecipitáció
CV	Controlled Vocabulary	felügyelt szókincs
CSAH	Charged single α-helix	magányos α-hélixek
DIP	Database of Interacting Proteins	-
DM	Disease-Causing Germline Mutation	betegséget okozó csíravonal mutáció
DNS	Deoxyribonucleic acid	dezoxiribonukleinsav
DO	DiseaseOntology	betegség ontológia
DOM	Domain	domén
ELM	Eukaryotic Linear Motif	-
FP	Fluorescence Polarization	fluoreszcens polarizáció
FPR/SPR	Surface Plasmon Resonance	felületi plazmon rezonancia
FRET	Förster or Fluorescence Resonance Energy Transfer	Förster-féle rezonancia-energiaátadás /fluoreszcens rezonáns energiaátadás
G2C	Genes2Cognition	-
gDa	gigaDalton	gigaDalton
GK	Guanylate Kinase	guanilát kináz
GO	GeneOntology	gén ontológia
H	Hydrophobe	hidrofób
H-híd	Hydrogen	hidrogén

DOI:10.15774/PPKE.ITK.2023.001

Rövidítés	Angol megfelelő	Magyar megfelelő
HMM	Hidden Markov models	rejtett Markov modell
hpa	Human Protein Atlas	humán fehérje atlasz
HTP	Human Transmembrane Protein	humán transzmembrán fehérje
HUPO	Human Proteome Organization	humán fehérje szervezetet
I2D	Interologous Interaction Database	-
IDP	Inrinsically Disordered Protein	rendezetlen fehérje
IDR	Intrinsically Disordered Region	rendezetlen régió
IMEX	International Molecular Exchange Consortium	-
ITC	Isothermal titration calorimetry	izotermális titrációs kalorimetria
Krio-EM/Crio-EM	Cryogenic Electron Microscopy	kriogén elektron mikroszkópia
LC-MS/MS	Liquid Chromatography with tandem mass spectrometry	folyadék kromatográfia és tömeg spektrometria
LIG	Ligand	ligandum
LLPS	Liquid-Liquid Phase Separation	folyadék-folyadék fázis szeparáció
LTD	Long Term Depression	hosszútávú depresszió
LTP	Long Term Potentiation	hosszútávú potenciáció
MAGUK	Membrane-Associated Guanylate Kinases	membrán-asszociált guanilát kináz
MI	Molecular Interaction	molekuláris interakció
MIMIx	Minimum Information about a Molecular Interaction eXperiment	minimális információ a molekuláris interakciós kísérletekről
mRNS	messengerRNA	hírvivő RNS
MS	Mass Spectrometry	tömegspektrometria
NMDAR	N-methyl-D-aspartate receptor	N-metil-D-aszparaginát receptor
NMR	Nuclear Magnetic Resonance	mágneses magrezonancia
nNOS	Neural Nitric Oxide Synthase	neurális nitrogén-monoxid-szintáz
OLS	Ontology Lookup Service	-
OMA	Orthologous MAtrix	-
OR	Odds ratio	esélyhányados

DOI:10.15774/PPKE.ITK.2023.001

Rövidítés	Angol megfelelő	Magyar megfelelő
P	Polar	poláris
PDB	Protein Data Bank	-
PFAM	Protein Family	-
PIK3	Phosphoinositide-3-Kinase	a foszfatidil-inozitol 3-kináz (PIK3)
PIK3R2	Phosphoinositide-3-Kinase Regulatory Subunit 2	A foszfatidil-inozitol-3-kináz szabályozó alegység
PISA	Protein Interfaces, Surfaces and Assemblies	-
PM	polymorphism	polimorfizmus
PPI	Protein-Protein Interaction	fehérje-fehéjre interakció
PPIDM	Protein-Protein Interaction Domain Miner	-
PS	Postsynapse	posztszinapszis
PSD	Postsynaptic Density	posztszinaptikus denzitás
PSD-95	Postsynaptic Density Protein 95	posztszinaptikus denzitás fehérje-95
PSI	Proteomics Standards Initiative	-
PSINDB	Postsynaptic Interaction Database	-
PSSM	Position-Specific Scoring Matrix	pozícióspecifikus pontozó mátrix
PTM	Post Translational Modification	poszttranszlációs módosítás
RNS	Ribonucleic acid	ribonukleinsav
RSA	Relative Surface Accessible Surface Area	relatív hozzáférhető felület
SCOP	Structural Classification of Proteins	fehérjék szerkezeti osztályozása
SLIM	Short Linear Motif	rövid lineáris motívum
SNV	Single Nucleotid Variation	egy pontos varáció
SPR	Surface Plasmon Resonance	felületi plazmon rezonancia
SynGAP	Synaptic Ras GTPase-activating Protein 1	szinaptikus RAS GTPáz aktiváló fehérje 1
SynGO	Synaptic Gene Ontologies	-
TARP	TCR Gamma Alternate Reading Frame Protein	-
TM	Transmembrane	transzmembrán
X-ray	X-ray crystallography	röntgen krisztallográfia
Y2H	Yeast Two-Hybrid System	élesztő két hibrid rendszer

DOI:10.15774/PPKE.ITK.2023.001

Ábrajegyzék

Ábrák

DOI:10.15774/PPKE.ITK.2023.001

Táblázatok

Függelék

DOI:10.15774/PPKE.ITK.2023.001

Absztrakt

Agyunkban milliárdnyi neuronális kapcsolat alakul ki idegsejtek között. Két idegsejt kapcsolatát szinapszisnak nevezzük, amelynek egyik funkcionális egysége a posztszinaptikus neuron membrán alatti jelfeldolgozó apparátusa, melyet posztszinaptikus denzitásak (PSD) nevezünk. Ezt mintegy 2 000 egyedi fehérje 10 000 molekulányi, összesen nagyságrendben 1 gigaDalton (gDa) tömegű rendszere alkotja. Ezen hálózaton belül a fehérje-fehérje kölcsönhatások eredményeként számos komplex képződik, amelyek összetétele nem csupán dinamikusan változik, de jellegzetes eltéréseket is mutat az egyes agyi régiók, illetve idegsejttípusok között. A komplexek felépítésére egyre több kísérletes információ áll rendelkezésünkre, azonban ismereteink legnagyobb része az itt található fehérjék bináris kapcsolataiból áll, a magasabbrendű komplexekről és különösen a teljes háromdimenziós hálózat szerveződéséről igen keveset tudunk. A posztszinapszist (PS) olyan alapvető neuronális folyamatokhoz kötötték már, mint a tanulás és a memória. Emellett alapvető szerepét az is jól jelzi, hogy több, mint száz neuronális betegséget lehet az itt található fehérjékkel összefüggésbe hozni. A doktori munkám fókuszpontjában a posztszinapszis rendszerének jobb megismerése állt a fehérjéin keresztül, elsősorban mutációs, fehérje szerkezeti és interakciós megközelítéssel, *in silico* módszerek segítségével. Ehhez a területen alkalmazott korszerű bioinformatikai eljárásokat használtam. Mivel a posztszinapszis pontos fehérje összetétele a fentebb említett sajátosságok miatt ma még nem ismert kellő pontossággal, első lépésként megbízható adatokat integrálva meg kellett határoznom a posztszinaptikus fehérjék listáját. Az így kapott fehérjék szerkezeti egységeit és ezeknek mutációk és polimorfizmusok általi érintettségét vizsgáltam. A tudományos konszenzussal megegyezően a transzmembrán régiók és domének betegséget okozó mutációknak való nagyobb kitettségét, míg a rendezetlen régiók alacsony érintettségét tapasztaltam. Fontos, és nem várt eredményként ugyanakkor a coiled-coil régiók esetében a mutációk alacsonyabb előfordulását találtam. A jelenség alacsony irodalmi feldolgozottsága miatt érdemes volt részletesen vizsgálnom a coiled-coil mutációkat szekvenciális, szerkezeti, funkcionális és betegségekkel való összefüggéseik szempontból, nem csak a posztszinapszis fehérjéin, hanem a teljes proteomon. Ennek eredményeként kimutattam a coiled-coilok N-terminális régiójában a betegséget okozó mutációk halmozódását, ami egyértelműen mutatja ezen régió fontosságát. Kimutattam azt is, hogy a coiled-coil mutációk nagy részben a központi idegrendszeri betegségekhez járulnak hozzá. A doktori munkám második felében visszatértem a posztszinapszis vizsgálatához és felépítettem a PostSynapticInteractionDataBase (PSINDB) adatbázist, amely alapját az irodalomból gyűjtött, más adatbázisokban eddig nem szereplő interakciós adatok adják, amelyhez saját annotációs rendszert is meghatároztunk a tudományterület legjobb gyakorlatai alapján. Az így kapott adatokat kiegészítettük egyéb, az interakciós és a fehérje kapcsolatok szempontjából releváns adatokkal. Ezen információkat felhasználva végül elemeztem az ismert kötőrégiók szerkezeti sajátosságait, és eddig nem azonosított, ám a kötésben szerepet játszható régiókra tettem javaslatokat.

DOI:10.15774/PPKE.ITK.2023.001

Abstract

In the human brain, millions of connections are created by neurons, this specific contact is the synapsis. A crucial functional unit located beneath the postsynaptic membrane is called the postsynaptic density (PSD). 2 000 individual proteins are found here and around 10 000 molecules form a 1 megaDalton weight system altogether. PS is a network of proteins where a number of complexes are formed from binary interaction. The network can change dynamically and differs between brain regions and neural cell types. Growing experimental evidence accumulates about the complexes, although most information has been collected from binary protein-protein interactions. We have a lack of knowledge about higher order assemblies and especially about the complete three dimensional architecture of the PS. The PS has an important role in fundamental neuronal processes such as learning and memory. This influential role is also highlighted by the fact that more than a hundred neuronal diseases were linked to its protein. The primary focus of my thesis is on a better understanding of the PS system through its proteins mainly from a mutational, structural and interactional view using in silico methods. For this purpose, I used advanced technologies from the field. The first consideration was to define a postsynaptic protein set, because there is great uncertainty in the composition of the PS as the result of the above mentioned peculiarity of the organelle. In the first step, I integrated information from accurate sources and created a list of proteins. Using this data, I investigated the relation between structural characteristics of the proteins and mutations – both disease causing inheritable and polymorphisms as well. The results for transmembrane and disordered segments were in line with other scientific results as the former are more and the later are less vulnerable to disease mutations. Interesting and not expected result was that coiled-coils are not accumulating disease mutations. This phenomenon is not well described in the literature, therefore I examined it in more detail. I studied it at the sequential, structural and functional level and for the whole proteome, not just for proteins of PS. Resulting data showed that disease-causing mutations accumulate at the N-terminal regions of coiled-coil, suggesting an essential role in the coiled-coil structure. Another conclusion was that coiled-coil mutations are mainly linked to central nervous system diseases. For the second half of my PhD, I returned to the investigation of the postsynapsis and set up the PSINDB database by collecting interactions from the literature that were not included in the primary interaction databases then. I defined an annotation system based on the best practices of the field. The collected interaction data were completed with more interaction data derived from other resources. Moreover, the data were supplemented by additional information that are important regarding protein interactions. Finally, using the collected data, I analyzed binding regions and suggested ones that are not included in PSINDB.

DOI:10.15774/PPKE.ITK.2023.001

1 Bevezetés

Agyunk milliárdnyi neuronból álló hálózat, az emberi test legkomplexebb szerve. Agyunk rendszere egy folyamatosan változó, dinamikus hálózat, melyben napról-napra kapcsolatok keletkeznek és szűnnek meg. A kialakult kapcsolatok felelősek az emberi érzelmek, gondolatok és viselkedés kialakulásáért. Nagy valószínűség szerint az emberiség történetével egyidős lehet a vágy agyunk működésének megismerésére. Az idegrendszer tanulmányozására utaló írásos bizonyítékok már időszámításunk előttről, az ókori Egyiptomból is rendelkezésre állnak. Az emberiség fejlődése során egyre több és egyre kisebb egységek váltak megfigyelhetővé. Míg az egyiptomiak az agy egészét, a 19. században Cayal már az idegsejteket vizsgálta mikroszkóp alatt, míg mára egyetlen neurális sejtben kifejeződő géneket, sőt, akár az általuk kódolt fehérjék elhelyezkedését is képesek vagyunk meghatározni. Az eltelt közel 4 000 évben egyetemes tudásunk elképesztő mértékben gyarapodott, azonban agyunk működésének teljes megértéséhez még messze nem érkeztünk el. Az agyunkat felépítő kommunikációs egységek a szinapszisok, amelyről egyre összetettebb képet látunk. Korábban molekuláris kapcsolóként tekintettünk rá, ma azonban már tudjuk, hogy a kapcsolatokban bonyolult "számításokat" végző rendszerről van szó. A szinapszis kialakításában részt vesznek a preszinaptikus és posztszinaptikus neuronok, előbbiek az információ leadásáért, utóbbiak az információ fogadásáért felelősek. A posztszinaptikus oldal - amely a dolgozatom tárgyát képezi - egy több ezer fehérjéből álló hálózat, amely olyan alapvető funkciókért felelős, mint a memória molekuláris szintű mechanizmusai, például a hosszú távú potenciáció (long term potentiation, LTP). A posztszinapszis működésének megértését nem csupán alapvető folyamatainak megismerése vezérli. Számos betegség, mint az Alzheimer, Parkinson, Huntington és a skizofrénia esetében kimutatták már a posztszinapszis fehérjéinek érintettségét. A posztszinapszis fehérjéinek vizsgálata több évtizede tart. Bioinformatikai módszerekkel nagyskálásan eddig a posztszinaptikus fehérjéket kódoló gének evolúcióját vizsgálták, illetve több tanulmány született, amelyben meghatározott fehérjék interakciós hálózatát térképezték fel. A mi kutatócsoportunk is végzett már korábban nagyskálás elemzést, amelyben a posztszinaptikus fehérjék szerkezeti sajátosságait térképezték fel. Erre alapozva végeztem a saját önálló kutatásomat, amelyben a posztszinaptikus fehérjék szerkezetének és az azokban előforduló mutációk összefüggését vizsgáltam. Emellett egy dedikált posztszinaptikus adatbázist is felállítottam, amely szintén szerkezeti aspektusból világít rá a fehérjék közötti interakciós lehetőségekre. Kitekintésként a humán proteom coiled-coil régióinak mutációs mintázatát is felmértem.

DOI:10.15774/PPKE.ITK.2023.001

2 Irodalami áttekintés

2.1 A posztszinapszis

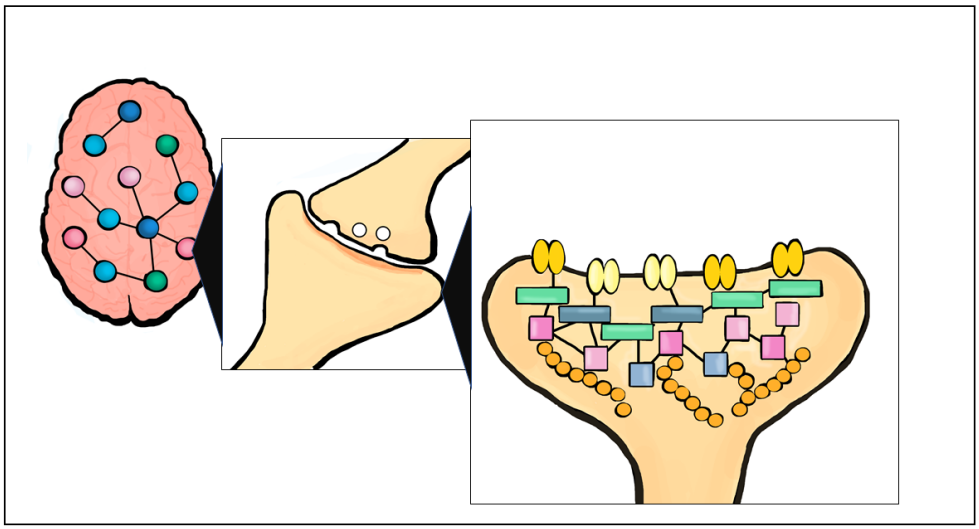
2.1.1 Szinaptikus jelátvitel alapjai, a posztszinapszis, mint meghatározó szereplő

Az agy a legkomplexebb szervünk, összetett funkciójáért az egymással speciális kapcsolatot kialakítani képes idegsejtek felelősek. Az agy hálózatát milliárdnyi neuron billiárdnyi (több, mint 10^{15} számú) szinapszisa hozza létre, ahol egy-egy neuron ∼7 000 szinapszison keresztül kommunikál más neuronokkal [1, 2, 3]. A szinaptikus jelátvitel kémiai, illetve elektromos jelek révén valósul meg. Előbbi esetében neurotranszmitterek, utóbbiak esetében kis molekulák (például ionok) közvetítik az információt. A jelátvitel jellemzői és feladatai nem csak a hírvívő molekulákban térnek el: az elektromos szinapszisok egyszerűbb felépítésűek és gyors válaszadásra teszik alkalmassá a kapcsolódó neuronokat, míg a kémiai szinapszisok változatosabbak és összetettebbek, ebből adódóan pedig komplexebb viselkedési mintázatok kialakulásáért felelősek. A központi idegrendszerben a tisztán kémiai szinapszisok legelterjedtebbek, azonban ma már kimutattak vegyes (kémiai és elektromos) szinapszisokat is. Az emlősök központi idegrendszerében a kémiai szinapszisok funkciója és a felszabaduló neurotranszmitterek alapján serkentő, illetve gátló szinapszisokat lehet elkülöníteni [4]. A klasszikus elképzelés szerint a szinapszisok csupán fizikai kapcsolóként működnek egyes neuronok között, ma azonban már tudjuk, hogy ennél bonyolultabb a szerepük, az információ feldolgozására és a jelek mintázatának felismerésére is alkalmasak (1. ábra) [5]. A kémiai szinapszisok felépítése jellegzetes elemekből tevődik össze: a preszinaptikus terminált és a posztszinaptikus neuront egy 15-20 nm-es szinaptikus rés választja el. A preszinaptikus neuron esetében általában axonok alakítják ki a kapcsolatokat, de dendritek is részt vehetnek benne. A posztszinaptikus terminált elsősorban dendrittüskéken találjuk [6]. A szinaptikus jelátvitel két fő egymást követő lépésből áll: egy adó, illetve egy vevő oldali folyamatból. A preszinaptikus neuronban az akciós potenciál hatására megvalósul a neurotranszmitterek felszabadítása, a posztszinaptikus oldalon pedig a receptorok segítségével jel fogadása. Míg a preszinaptikus oldalon kanonikusan, azaz gyakorlatilag ugyanazon fő mechanizmussal és molekulák részvételével valósul meg a jel leadása, addig a posztszinaptikus oldalon diverz receptorok képesek a jel fogadására. Mind a preszinapszis, mind a posztszinapszis esetében bonyolult fehérjehálózatok vesznek részt az információ továbbításában [7]. A szinaptikus információátadás összetettségéhez hozzájárul, hogy a szinapszis felépítése jelentősen módosul a különböző ingerek, folyamatok (pl. tanulás) során és a szinapszis felépítése sejt, inter- és intramolekuláris szinten is dinamikusan változik [3].

2.1.2 A posztszinapszis jellemzői, a posztszinaptikus denzitás

Az elmúlt évtizedekben számos nagyskálás kísérletet végeztek a PS fehérje-összetételének meghatározására, elsődlegesen a PSD fehérjéinek azonosítására fókuszálva. Ezen vizsgálatok eredményei eltérő számú

DOI:10.15774/PPKE.ITK.2023.001



1. ábra A központi idegrendszer szerveződési szintjei és komplexitása Bal: a különböző agyi régiók eltérő fehérje összetételt mutatnak. Középen: a szinapszisok axon-axon és dendrit-dendrit között is létrejöhetnek. Jobb: a posztszinapszis összetett fehérje hálózata

fehérjét mutattak ki: összesen ∼2 000 fehérjére mondhatjuk, hogy a tágan értelmezett PS proteom része, amiből ∼800 fehérje esetén van egyetértés az eredmények között [8]. A PS proteomban a fehérjék elenyésző százaléka található monomer formában, a nagy részük komplexekbe szerveződik (ld. 2.3.2. fejezet). Eddig több, mint 220 szupramolekuláris komplexet azonosítottak, amelyek olyan makromolekuláris összeszerelődések [5], amelyek több komplexből épülnek fel. Az elmúlt évtized tudományos eredményei (ld. krio-elektron mikroszkóp, 2.2.3. fejezet) nagyban felgyorsították a komplex szerkezetek vizsgálatát, azonban a feltételezett komplexeknek csak kis hányadát ismerjük csak [9]. A posztszinaptikus neuronon belül jól elkülöníthető a posztszinaptikus denzitás, mely egy elektronmikroszkóppal is megfigyelhető, nagy sűrűségű, félig membránhoz kötött struktúra, jellegzetes felismerhető képét a fehérjék nagy koncentrációja adja [3]. A PSD-ben legalább 800 különböző fehérje található meg [10]. Egy PSD átlagos tömege 1 gDa, ami körülbelül 10 000 fehérjemolekulának feleltethető meg [6]. Az egyes fehérjék koncentrációja is jelentős eltérést mutat, a leggyakrabban előforduló fehérjék a PSD-95 (∼300 kópia), a SynGAP (∼360 kópia), a CAMKII (∼800-4 000 kópia) [6, 11]. A neuronon belül az eltérő fehérje expressziós szintek mellett fontos megemlíteni, hogy a különböző agyi területek neuronjaiban található PSD-k heterogenitása is jelentősen eltér. Például a neocortex és a hippokampusz részekben megtalálható neuronok PSD-je jelentős diverzitást mutat a fehérjék összetételét tekintve, míg az agytörzs a skála ellentétes pólusán található. Marker fehérjék segítségével egerekben kimutatták, hogy a PSD összetétele az életciklus során is változást mutat [12]. A fehérjék előfordulása ezen felül különböző fajokban is mutat eltéréseket, például egerek esetében a több jelátviteli útvonallal is kap-

DOI:10.15774/PPKE.ITK.2023.001

csolatba lépő, nagy komplexekké összeszerelődni képes PSD scaffold fehérjék előfordulása gyakoribb, mint a humán minták esetében [10]. Érdekes viszont, hogy a szinaptikus fehérje családok előfordulása evolúciósan jelentősen konzervált, már az egysejtű organizmusokban is megtalálhatóak a funkcionális szempontból legmarkánsabban jellemző csoportok [5]. A serkentő szinapszisokban a PSD laminárisan erőteljesen struktúrált organizáció. Három fő részre osztható a szerkezete (1. ábra).

- ioncsatornák, receptorok és sejt adhéziós molekulák. A réteg feladata alapvetően a preszinaptikus információ fogadása.
- MAGUK állvány fehérjék és stargazin. Ez a réteg összeköti a receptorokat a PSD többi részével, egy "állványzatot" létrehozva, aminek segítségével más fehérjék fizikailag képesek közel kerülni egymáshoz.
- enzimek, másodlagos scaffodok, citoszkeletális fehérjék és azokkal asszociált fehérjék

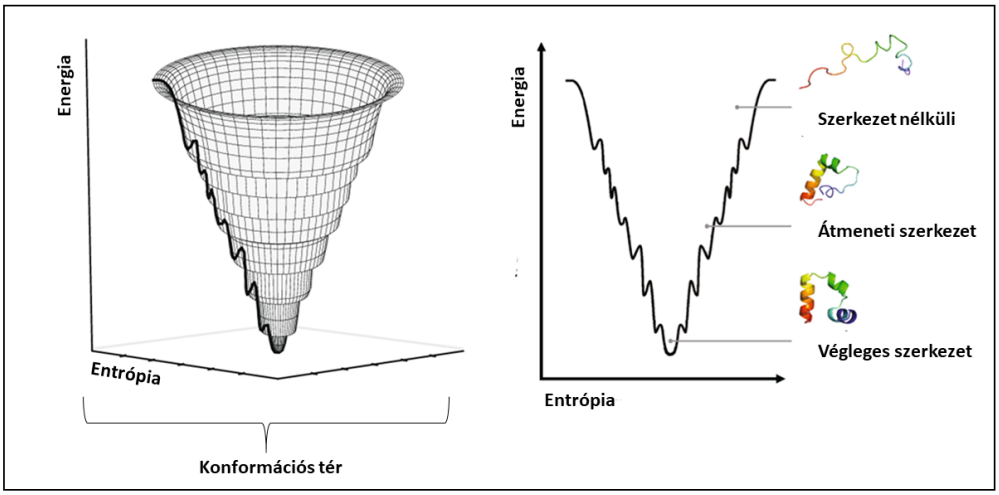
A gátló szinapszisok esetében is kimutattak már egy, a PSD-hez hasonló struktúrát [13]. Az utóbbi évek modern mikroszkópos eljárásainak fejlődése lehetővé tette, hogy a PSD planáris elrendeződését is megfigyeljék, ahol a fehérjék nanoklaszterekbe vagy más néven nanodoménekbe rendeződnek [1]. A PSD-t alkotó funkcionális egységek (komplexek) direkt kölcsönhatások révén vagy scaffold molekulákon keresztül megvalósított kapcsolódásokkal alakulnak ki. A funkcióval összefüggő megfelelő lokalizáció érdekében a komplexek membránhoz horganyozva vagy citoszkeletális elemekhez kötve is jelen lehetnek [6]. Funkció szempontjából a legnagyobb számban expresszált fehérjék a citoszkeletont alkotó fehérjék (12%), regulátor fehérjék és kinázok (11%), GTPázok (8%), sejtadhéziós molekulák (7%), metabolizmushoz köthető fehérjék (7%), receptorok és ioncsatornák (6%) és a már említett állvány fehérjék (6%) [6]. Az állványfehérjék különösen fontos szerepet töltenek be a PSD működésében. Jellemzően számos kölcsönhatás kialakítására alkalmas doménnel rendelkezhetnek, amelyek segítségével egyszerre több fehérjéhez kapcsolódhatnak [10, 14].

2.2 A fehérjék funkcionális és szerkezeti sajátosságai

2.2.1 A fehérjék feltekeredése

A fehérjék jelentik a fő végrehajtó elemeket a molekulák, a sejtek és az organizmusok szintjén is [15]. A legtöbb fehérje úgy alakult ki, hogy képes legyen a környezetében lévő többi molekulával (DNS-sel, RNS-sel, másik fehérjével stb.) kapcsolatba lépni [15]. A fehérjék egyszerű láncként szintetizálódnak a húszféle, természetben is előforduló aminosav egymás után fűződésével, azonban funkciójuk betöltéséhez szükséges megfelelő háromdimenziós szerkezet kialakulása [15, 16]. Anfinsen már a 20. század közepén bizonyította, hogy a szerkezet felvételéhez szükséges információt az aminosavak sorrendje önmagában hordozza és meghatározza, és ez a szerkezet pedig nélkülözhetetlen a fehérje funkciójának ellátásához [17]. Csupán az aminosavak egymásutánisága azonban hatalmas teret enged a lehetséges

DOI:10.15774/PPKE.ITK.2023.001



2. ábra A globuláris fehérjék feltekeredésének energetikai térképe [19]. A fehérje a feltekeredése során egy nagy energiájú állapotból jut el egy alacsonyabb energiájú állapotba. Bizonyos fehérjék esetén több alacsony energiájú végállapot is létezik (rendezetlen fehérjék).

fehérje konformációk kialakulásához. Az Anfinsen-dogma után nem sokkal született Levinthal-paradoxon (1969) szerint, ha egy 101 aminosav hosszú fehérjelánc minden aminosav párja 10 konformációban fordulhat elő, és minden konformáció felvétele 10-13 másodperc, akkor is egyetlen fehérje feltekeredéséhez hosszabb időre (10^{27} év) lenne szükség, mint amennyi ideje az egész univerzumunk létezik ($\sim$13 x10^{12} év) [18]. Ebből következően a fehérjék felgombolyodása ("folding", "feltekeredés") nem lehet egy teljesen véletlenszerű folyamat eredménye. Az elméletben szinte végtelen konformációs lehetőség közül a természetben azok fordulnak elő, amelyek termodinamikailag a leginkább stabilak. A termodinamikailag stabil állapotokra alacsony szabad energia és entrópia jellemző - ebben az állapotban van a fehérje a natív szerkezetében [16]. A különböző fehérjék esetében a natív konformáció felvétele eltérő módon valósul meg, sokszor köztes, lokálisan alacsonyabb energiaállapotú tranziens pontok érintésén keresztül, többféle útvonalon keresztül is lejátszódhat (2. ábra)[15, 16].

A fehérjék nagyobbik része meghatározott szerkezettel rendelkezik, azonban ez nem jelenti azt, hogy szerkezetük rigid, konformációs flexibilitás jellemzi őket. Emellett a natív és a szerkezet nélküli állapotuk közötti kis energia különbség lehetővé teszi a fehérjék szükséges flexibilitásának biztosítását. A fehérjék stabilitása (a natív konformáció megtartására való hajlandóság) nem egyforma az egész szerkezetben, stabilabb és kevésbé stabil szakaszok felváltva is előfordulhatnak egy-egy fehérjén belül [15]. Emellett a fehérje feltekeredési egységek a fehérje élettartama során fel- és letekerednek, ebből következően a fehérjék nem kétállapotúak, hanem többállapotúak. A feltekeredési folyamat azonban egy bonyolult, és akár veszélyes folyamat [20]. Habár a nem megfelelően feltekeredett fehérjék továbbra is funkcióképesek lehetnek, ám megbonthatják a sejtek normális működését, például aggregátumokat

DOI:10.15774/PPKE.ITK.2023.001

képezve felhalmozódhatnak. A nem megfelelő fehérje feltekeredés több száz betegség kiváltója lehet [21], különösen gyakori neuronális megbetegedések során (ld. 2.4.3 fejezet), pl. a Lewey-testes demencia esetében az MRI-felvételeken sokszor szabad szemmel is láthatóak az ún. Lewey-féle zárványok, melyek alfa-szinuklein és ubikvitin aggregátumok.

A fehérje feltekeredési probléma megoldása a modern molekuláris biológia egyik legnagyobb kihívása. Számos kísérletes (ld. 2.2.3. fejezet és 2.2.4. fejezet) és számítógépes módszer létezik a fehérje szerkezetek meghatározására. Az elmúlt években több tudományos fejlesztés segítette, hogy közelebb kerüljünk a probléma megoldásához, azonban fontos kiemelni, hogy ezek mindig egy adott fehérje (komplex) szerkezetének megoldását jelentik, a feltekeredés általános biofizikai törvényeken alapuló leírása nélkül. Kísérletes területen a krio-elektronmikroszkópia (Kémiai Nobel-díj, 2017) jelent áttörést a nagyobb fehérje komplexek meghatározására (ld. 2.2.3. fejezet), míg számításos területen az AlphaFold2 (2020), megjelenése jelentett egy-egy mérföldkövet (ld. 2.2.4. fejezet)[22].

2.2.2 A fehérjék szerkezete

A fehérjék szerkezeti szempontból többféle szinten definiálhatók, és a különböző szerveződési szinteket más stabilizáló erők befolyásolják: A polipeptid láncon belül az egymást követő aminosavak között kialakuló kovalens kötések határozzák meg a fehérjék elsődleges szerkezetét. Ezt a kapcsolódást peptidkötésnek nevezzük, ilyenkor egy aminosav karboxilcsoportja és a szomszédos aminosav aminocsoportja lép kölcsönhatásba, miközben víz lép ki (ebből fakadóan nevezzük a fehérjében levő aminosavakat reziduumnak is, mivel a kölcsönhatás után az eredeti aminosav egy darabja szerepel már csak a fehérjében). A másodlagos szerkezet az aminosavak közötti hidrogén-híd kötések eredményeként létrejött lokálisan stabil szerkezet. Alapvetően két prominens osztálya létezik az α-hélix és a β-redő [16]. A lokálisan stabil másodlagos szerkezetek kapcsolatba léphetnek egymással, így létrehozva a fehérjék harmadlagos szerkezetét, amit a hidrofób kölcsönhatások vezérelnek. A hidrofób kölcsönhatás a legdominánsabb erő a fehérjék feltekeredésének szempontjából (még ha önmagában relatív gyenge kölcsönhatás is). A harmadlagos szerkezet kialakulását segítik továbbá elektrosztatikus és Van der Waals-féle erők is. A fehérjék negyedleges szerkezete az a natív konformáció, amelyben a szerkezet ellátja a biológiai funkcióját. Számos fehérje esetében ez azt jelenti, hogy több alegységből álló oligomerek jönnek létre, azonos, vagy más fehérjékkel együttműködve [15]. A fehérjeszerkezeteket egy olyan polipeptid láncként is leírhatjuk, ahol a másodlagos szerkezeti elemek követik egymást, amiket az egyik csoportba (α, β) sem sorolható random coil szegmensek kötnek össze. Ezek bizonyos mintázatot mutatnak, amelyek visszatérnek a láncban, vagy akár különböző fehérjékben. Ezek a mintázatok szupermásodlagos szerkezeteket hoznak létre, ahol legalább két másodlagos szerkezeti elem összekapcsolódik (pl. β-α-β-loop)[16]. A fehérjék szerkezetének a fenti kategóriákba csak részben illeszthető, de fontos egységei az úgynevezett domének. A domén definíciója nem egységes, a dolgozatban úgy te-

DOI:10.15774/PPKE.ITK.2023.001

kintettem, hogy ezek önálló feltekeredésre (más szóval foldingra) képes, autonóm, funkcióval rendelkező fehérje egységek. A domén a fehérjék építőkockái: a legegyszerűbb fehérjék csupán egyetlen doménből állnak, de a domének ismétlődhetnek a fehérjén belül, és különböző fehérjékben is megjelenhet ugyanaz a domén [6].

2.2.2.1 A coiled-coil szerkezeti elem

A coiled-coilok az elsők között felfedezett szupermásodlagos szerkezeti elemek, ahol két vagy több α-hélix alkot spirált egymással úgy, hogy maximalizálják a hidrofób kontaktusok kialakulását a két lánc között [23]. A különböző organizmusokban gyakran fordulnak elő, általában a különböző proteomok 3-10%-át teszik ki [24, 25]. 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 [23]. Az egyes pozíciókat a heptádban meghatározott karakterekkel jelölik [26, 27] , 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’ (3. ábra). A Peptid Velcro (1993) hipotézis szerint a coiled-coil szerkezet kialakulásához három feltételnek kell teljesülnie

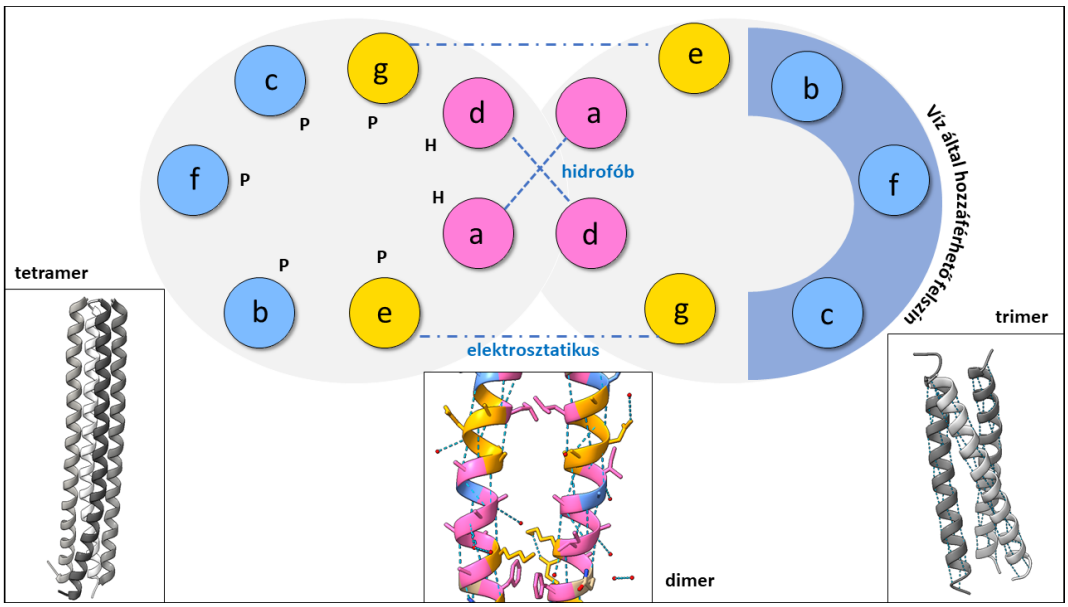
1. Az ‘a’ és ‘d’ pozíciókban hidrofób aminosavak, amelyek hidrofób és van der Waals kötések révén stabilizálják a szerkezetet

2. Az ‘e’ és ‘g’ pozíciókban töltött aminosavak helyezkednek el, amelyek hélixek közötti elektrosztatikus kölcsönhatásokat biztosítják többi pozícióban pedig hidrofil aminosavaknak kell lenniük, mivel a coiled-coil szerkezet ezen része van kitéve a vízzel való érintkezésnek [24].

A coiled-coilban található ismétlések száma kettőtől - extrém esetben a kétszázig is terjedhet [24], a természetben a kettő és hat közötti heptád a legelterjedtebb [25]. Azt feltételezik, hogy a szerkezet kialakulásában a fentieken kívül kulcsszerepe lehet egy trigger szekvenciának nevezett szakasznak is. Ez a rövid szekvenciarészlet kódolhat olyan önálló feltekeredésre képes egységet, ami leszűkíti a feltekeredés terét és coiled-coil szerkezet felvételéhez segíti a fehérje következő részét [24]. A coiled-coilok között a leggyakoribbak a két vagy három hélixből kialakuló szerkezetek, de magasabb szintű összerendeződés is lehetséges [28]. Attól függően, hogy hány hélix kapcsolódik a coiled-coilban, megkülönböztethetünk különböző oligomerizációs állapotokat. Az oligomerizációs állapotot ‘a’, ‘d’, ‘e’ és ‘g’ pozícióban lévő aminosavak határozzák meg. A H/P mintázatot megváltoztatva különböző oligomerizációs állapotok érhetőek el [27]. Például, ha az ‘a’ pozícióban izoleucin és a ‘d’ pozícióban leucin van, akkor egy dimer fog kialakulni, ennek ellenkezője tetramer coiled-coil kialakulásához vezet [29].

A láncok minden oligomerizációs állapot esetében parallel és antiparallel orientációban is előfordulhatnak [26]. A láncok származhatnak azonos fehérjékből is, ilyen esetben homooligomerekről beszélhetünk, vagy különböző fehérje láncokból, amiket heterooligomereknek nevezünk [25]. Bizonyos esetekben a

DOI:10.15774/PPKE.ITK.2023.001



3. ábra A coiled-coilok szerkezeti felépítése Felül: A coiled-coil alapja a heptád ismétlés ('abcdefg') ahol HPPHPPP aminosavak követik egymást (H: hidrofób, P: poláris) Az 'a' és 'd' pozícióban hidrofób aminosavak (rózsaszín), a 'g' és 'e' poziciokban töltött (sárga) többi pozícióban hidrofilek (kék) A coiled-coil magját a hidrofób és elektrosztatikus kölcsönhatások tartják össze (magát a hélixet a hidrogénhidak) Alul: A coiled-coil-ok különböző oligomerizációs állapotokban fordulhatnak elő - dimer, trimer, tetramer stb.

coiled-coil egy fehérjén belüli két hélixből is összeállhat. Léteznek úgynevezett nem kanonikus coiled-coil formák is. Ezekben az esetekben a hélix alapját nem a heptád-ismétlés képezi [23]. *Giardia lamblia* esetében felfedeztek olyan citoszkeleton asszociált fehérjéket, amelyekben a heptád ismétlődések közé beékelődnek nem kanonikus formák (24, 25 és 26 aminosavat tartalmazó ismétlődésekkel) is, ami hosszabb szerkezeti motívumok kialakulásához vezetett [30]. Nagy általánosságban azonban elmondható, hogy a coiled-coilok hossza és az elsődleges szerkezete jelentősen konzervált [25].

A coiled-coilok szerkezeti diverzitása, változatos funkciók megjelenésében érhető tetten [29]. Gyakran előfordulnak motorfehérjékben, transzkripciós faktorokban és receptorokban is. Az egyik legjobban ismert, jól karakterizált funkciója a leucin cipzár létrehozása, ami számos fontos DNS-kötő fehérjében megtalálható [29]. Ilyenkor a hidrofób kölcsönhatást elősegítő leggyakoribb aminosav mindkét láncban a leucin, amelyek cipzárszerűen tartják össze a kettős spirált. A coiled-coilok hossza nagyon változó. A hosszabb coiled-coilok rigid szálakat hoznak létre, mint például a keratin vagy miozin, ami lehetőséget ad olyan felületek létrehozására, ami nagyobb szerkezeti komplexek kialakulását szolgálja, mint például az extracelluláris mátrix vagy a citoszkeleton hálózata [29]. Távoli régiók összekapcsolásában is szerepet játszanak, amikor molekuláris távtartónak is szokták őket nevezni (előfordulnak lipoproteinekben vagy például a kinetokorok esetében) [25]. Egy másik érdekes coiled-coil funkció, hogy molekuláris vonalzóként működnek a sejteken belül [25]. A rövidebb coiled-coilokat tartalmazó fehérjék jóval flexibilisebbek és olyan funkciókat látnak el, mint az oligomerizáció elősegítése, szignáltranszukció

DOI:10.15774/PPKE.ITK.2023.001

vagy kis molekulák transzportja [26].

2.2.2.2 Rendezetlen fehérjék

A tradicionális szerkezeti biológia számára “láthatatlan” régiók teszik ki a humán proteom nagyjából harmadát [31], amelyek jól meghatározható, 3D szerkezet nélküli szakaszok. Ezeket az úgynevezett rendezetlen részeket diverz konformációs állapotok jellemzik, mivel nincsen egyetlen jól meghatározott alacsony energiájú állapotuk, ellentétben a rendezett fehérjékkel (2. ábra). Ezen régiók szekvenciális jellemzői, hogy alacsony hidrofóbicitású, ugyanakkor magas nettó töltöttségű aminosavakból állnak [32]. Korábban ezekre a régiókra funkció nélküli távtartóként tekintettek [33], azonban az elmúlt húsz évben egyre több funkciót társítottak hozzájuk, ami szekvencia-szerkezet-funkció paradigma felülbírálatát eredményezte [34]. Szerkezeti szempontból a fehérjéket teljesen rendezett, rendezetlen régiókat tartalmazó (Intrisnically Disordered Region, IDR) és teljesen rendezetlenként (Intrisically Disordered Protein, IDP) írhatjuk le. Egy friss tanulmány szerint a humán proteom 49%-a teljesen rendezett, a fehérjék 19%-ban vannak IDR-ek és a fehérjék 32%-a IDP [32]. Ezen régiók változékonysága abban is megmutatkozik, hogy képesek akár rendezetlen-rendezett tranzícióra is [33], illetve bizonyos állapotok közötti gyors konverzióra is, amely a fehérje szakasz hozzáférhetőségét befolyásolja [31]. A szerkezetbeli lazább megkötések miatt gyorsabban evolválódnak, mint a globuláris domének [33]. A rendezetlen, illetve IDR-eket tartalmazó fehérjék többféle osztályozása létezik szerkezeti, illetve funkcionális hasonlóságok alapján. A következőkben egy fontos csoportot emelnék ki, ahol fontos szerepe van az előbb említett szerkezeti jellemzőknek.

2.2.2.2.1 Lineáris motívumok

A rövid lineáris motívumok (Short Linear Motif, SLiM) 3-10 hosszúságú szakaszok. Fontos funkcionális egységek a fehérjéken belül, ahol a funkció betöltéséhez egy partner doménnel való kölcsönhatás szükséges [33]. Sok esetben több kópiában fordulnak elő az adott fehérjében [35]. A SLiMek legtöbbször IDR-en belül alakulnak ki, és erősen konzerváltak. Ez a tulajdonságuk, ami legjobban segíti az azonosításukat, amíg a rendezetlen rész evolúciós szempontból igen változékony (ld. 2.2.2.2. fejezet), a benne levő pár aminosavnyi motívum szigetszerűen kiemelkedik amikor a konzerváltságot vizsgáljuk. Rövidségük alkalmassá teszi őket, hogy egy rendezetlen részben *de novo* képesek mutálódni. Mivel rövidek, ha változik a partner domén, könnyen vele változnak, képesek a konvergens evolúcióra. Alacsony affinitás és tranziens kölcsönhatás kialakítása jellemző rájuk. Ezt a tulajdonságot tudják kihasználni különböző patogének, amik mintegy “lemásolják” a motívumot és egy magasabb affinitású verzióval felülkompetálják a sejten belüli kölcsönhatásokat. Számos motívum-domén párt már a posztszinapszis esetében is azonosítottak. Ilyenek a foszfotirozin motívumok kölcsönhatása az SH2 doménekkel, C-terminális motívum - PDZ domén párok vagy a PxxP motívumok kapcsolódása SH3 doménekhez [33].

DOI:10.15774/PPKE.ITK.2023.001

2.2.3 A fehérjék szerkezetének és funkciójának összefüggése

A molekuláris biológia egyik legfontosabb célja a fehérjék funkciójának megismerése. A fehérje funkció megismeréséhez szükséges, de nem elégséges feltétel a fehérje szerkezetének ismerete, mivel a szerkezeti információ csak bizonyos szintig nyújt információt a fehérjék sejten belüli szerepéről [36]. Emellett, mint ahogyan azt az előző fejezetben is bemutattam, nem csak a szerkezet, hanem a „szerkezetnélküliség" is rendkívül fontos funkciókkal bír. Az IDR-ekre vagy IDP-kre szoktak „dark proteom"-ként (láthatatlan proteomként) is hivatkozni, amely elnevezés onnan származik, hogy a klasszikus szerkezetmeghatározási technika (röntgenkrisztallográfia) számára nem láthatóak ezek a régiók/fehérjék [37]. A röntgenkrisztallográfia mellett más módszerek is rendelkezésre állnak a fehérjék szerkezetének meghatározására. Ezek részletes ismertetésétől a dolgozatom terjedelmi korlátai miatt eltekintek, azonban a három meghatározó nagy felbontó képességű (atomi szint) eljárást előnyeikkel és limitációikkal áttekintem a 1. táblázatban (ezen technikákat megemlítem még a 3.3. fejezetben az interakciók meghatározási lehetőségeinél is). A fehérjék szerkezetének meghatározása bármely kísérletes technika esetén is egy rendkívül bonyolult, többlépcsős, erőforrásigényes feladat, és kivétel nélkül igen költséges műszereket igényelnek. Ezek a kísérletek azonban nem kiválthatóak más technikákkal az általuk feltárt szerkezeti információ részletessége miatt.

A fehérjék szerkezet-funkció összefüggéseinek ismerete további elemzésekre ad lehetőséget. Általánosságban elmondható, hogy a hasonló szerkezetű fehérjék hasonló funkciókat látnak el, például számos membránpórust képező fehérjében találunk β-hordó szerkezetet [38]. A szerkezeti hasonlóságot lokális, illetve globális szinten vizsgálhatjuk két fehérje között. Lokális hasonlóság vizsgálatánál a fehérjéket domének vagy annál is kisebb egységek szintjén hasonlítjuk össze. Abban az esetben, ha itt találunk egyezést, pl. adott aminosavak egy bizonyos aktív helyre vagy partnerkötő régióra jellemző térbeli elrendeződésében, az hasonló funkcióra utalhat. Fontos, hogy az ilyen jellegű hasonlóság nem feltétlenül jelent evolúciós rokonságot, mivel a funkcionális helyek egymástól függetlenül is megjelenhetnek különböző ősöktől származó fehérjékben is. Globális szerkezeti hasonlóság esetén domének szintjén hasonlítjuk össze a fehérjéket, ami már biztosabban mutathat evolúciós hasonlóságot. Ez azonban nem feltétlenül fog nagymértékű szekvenciális hasonlóságban is megmutatkozni, mivel a fehérje foldok sokszor akkor is megmaradnak, amikor a szekvencia hasonlóság már nagymértékben megszűnik (azonban kivételek is ismertek) [36]. A szerkezet mellett az aminosavszekvencia - amely egyszerűbben hozzáférhető - is számos információval szolgálhat, azonban ezek az adatok nem rendelkeznek hasonló robosztusággal, mivel a fehérje szerkezet evolúciósan jelentősen konzerváltabb, mint maga a szekvencia [39]. Ugyanakkor mivel a szekvenicális információ könnyebben hozzáférhető, számos bioinformatikai módszer alapja a hasonló szekvenciák megtalálása, és ez a módszer többnyire alkalmas rá, hogy megtaláljuk a hasonló szerkezetű (és funkciójú) fehérjéket is. Ez egy fontos lépés, hiszen a több százmillió ismert fehérje egyedi kísérletes vizsgálata több szempontból is lehetetlen feladat, de így az újonnan

DOI:10.15774/PPKE.ITK.2023.001

felfedezett fehérjéknek homológia alapján könnyebben megtalálhatjuk a funkcióját.

1. táblázat A legfontosabb szerkezetmeghatározó eljárások és azok előnyei és limitációi

Technika	Előny	Limitáció	Nobel-díj
Röntgenkrisztallográfia (X-ray)	Jól kidolgozott protokollok, megbízhatósági metrikák	Hiányzó rendezetlen régiók, nem natív környezet, kristályosítás sikere kérdéses	1962, May Perutz, John kendrew
Mágneses magrezonancia spektroszkópia (NMR)	Strukturális flexibilitás atomi felbontással, több időskálán vizsgálható	Méret limitáció, izotóp jelölés szükségessége, mintakészítés sikere kérdéses	2002, Kurt Wüthrich
Krio-elektronmikroszkópia (Krio-EM)	Natív állapot, nincs méret limitáció	A másik két módszerhez képest alacsonyabb felbontás	2017, Jacques Dubochet, Joachim Frank and Richard Henderson

2.2.4 Fehérjeszerkezet leírására használható számítógépes módszerek

A kísérletes módszerek mellett számos predikciós módszer áll rendelkezésre a fehérje szerkezetek becslésére, amelyek fontos eszköztárat jelentenek kísérletek megtervezéséhez, vagy nagyskálás elemzésekhez. Az első predikciós programok egyszerű statisztikai alapokon működtek, például a másodlagos szerkezetek (α-hélix, β-redő ld. 2.2.2. fejezet) meghatározásra és az egyes aminosavak előfordulásának valószínűsége alapján voltak képesek becsülni a szerkezeti elemek helyét a szekvencián belül. A későbbiekben bonyolultabb szerkezetek (például coiled-coil) és rendezetlen régiók becslésére is készültek programok. A jelenleg az alkalmazott módszereknek két nagy csoportja van: (1) tisztán a fehérjék fizikokémiai tulajdonságain és alapvető statisztikai módszereken alapuló eljárások, (2) a gépi tanuláson alapuló módszerek. Bár a gépi tanulás alapú módszerek egyfajta fekete doboznak tekinthetőek (azaz nem értjük pontosan, hogy milyen logikai döntések alapján hoznak döntést), a legtöbb területen mégis felülmúlják a klasszikus algoritmusokat, mivel a tisztán biofizikai és biokémiai leírása ezeknek a rendszereknek nem elég precíz. További gondot okoz a kísérletes módszerek korlátaival és az adatok esetleges pontatlan interpretációjával átkerülő zaj mennyisége. A 2000-es évektől a fehérjék háromdimenziós szerkezeti becslésére fejlesztett módszerek is egyre szélesebb körben kerültek felhasználásra. Ezeknek több kategóriája létezik az alapján, hogy globális vagy lokális mintázatokat használnak fel a szerkezetek felépítése során. Globális modellezési eljárásokat akkor alkalmaznak, amikor valamilyen rokon fehérje

DOI:10.15774/PPKE.ITK.2023.001

szerkezet rendelkezésre áll, de ide sorolhatóak azok a módszerek is, amelyek ko-evolúciós mintázatok alapján becsülnek térbeli kontaktusokat. Lokális modellezés esetében nem áll rendelkezésre rokon fehérje, amelynek teljes szerkezetét fel lehet használni a modellezés során. Ilyekor egyes fragmensek templát alapú modellezése és aztán a kapott darabok összeillesztése lehet a járható út. Az utóbbi években a szerkezetbecslő eljárásokban is egyre jobban elterjedtek a gépi tanulás alapú módszerek. 2021-ben a DeepMind által fejlesztett deep learning eljárás, az AlphaFold2 eredménye hatalmas előrelépést jelentett, mivel a korábbi ∼60%-ról 90% közelébe emelte a szerkezet becslés pontosságát monomer globuláris fehérjék esetén [22]. A módszer fragmens alapú valamint koevolúciós megközelítést kombinál, és többszörös szekvencia illesztéseket is felhasznál bemenetként. Mivel a fehérje feltekeredés és a kölcsönhatások kialakulása ugyanazon biofizikai szabályok alapján megy végbe, kiegészítő lépésekkel akár több fehérjéből álló, úgynevezett komplex szerkezeteken is nagy pontosságot képes elérni [40].

2.2.5 A poszttranszlációs módosítások

Az élő sejtekben számos molekula együttes hatása révén marad fenn a sejtek integritása és valósul meg a megfelelő funkciójuk. A fehérjék, mint a legfontosabb végrehajtó molekulák az aminosav sorrendjükben hordozzák biológiai funkciójukat, azonban számos olyan szabályozási mechanizmus létezik, amely finomhangolja pontos működésüket [41]. Ennek a finomhangolásnak köszönhetően a humán proteomban megtalálható körülbelül 20 000 fehérjéből gyakorlatilag több százezer különböző molekula jöhet létre [42].

Az egyik gyakori szabályozási mód az alternatív splicing, ami azt jelenti, hogy az mRNS érése során az exonok különböző kombinációkban maradhatnak bent a lefordítandó régióban, így a fehérje egy módosított szekvenciával jön létre. A kivágódó rész miatt a fehérje például másik sejt kompartmentbe juthat be, és ott fejti ki a hatását. Egy másik gyakori szabályozási mód a poszttranszlációs módosítás (Post Translational Modification, PTM), ami bizonyos esetekben visszafordítható folyamat [41] és egyes esetekben szorosan kötődik a lineáris motívumok speciális alosztályaihoz. A PTM-ek fontos szerepét hangsúlyozza az a megfigyelés, hogy majdnem minden fehérje esetében kimutattak legalább egy poszttranszlációs módosítást [43]. A PTM a transzláció után valósul meg, és leggyakrabban egyes aminosavak kémiai módosítását, pl. funkciós csoport hozzáadását jelenti. Azonban előfordulhat még az aminosav sorrend módosulása is pl. más peptidekkel vagy fehérjékkel való konjugáció vagy a fehérje splicing révén. Funkciójukat tekintve modulálják a fehérjék aktivitását, makromolekuláris interakcióikat, celluláris lokalizációjukat és számos fundamentális molekuláris folyamatban vesznek részt [41]. A fehérje kölcsönhatási preferenciáit egyrészről indirekt módon, a fehérje globális konformációs változásának előidézésével befolyásolhatják, másrészt sok esetben közvetlenül hatnak a molekuláris felismerésben részt vevő régiók elektrosztatikus vagy egyéb lokális strukturális jellemzőire [42].

A PTM-ek reverzibilis kovalens módosításának prototípusa az aminosavak foszforilációja. Ennek során

DOI:10.15774/PPKE.ITK.2023.001

a módosítás hatására egy többlet negatív töltés jelenik meg a fehérjén, ami befolyásolja, hogy az adott fehérje milyen más molekulákkal tud kölcsönhatásba lépni. Mivel a foszforiláció reverzibilis, ezt egy kapcsolónak is tekinthetjük, ami bizonyos kölcsönhatásokat ki-be tud kapcsolni. A posztszinapszisban nagyon fontos szerepet játszik a foszforiláció, amely kombinatorikus módon tudja finomhangolni a különböző interakciókat [43]. A foszforiláción túl azonban több, mint 200 [42], más források alapján több, mint 400 PTM ismert [43], amelyek közül a legfontosabb módosítások az 2. táblázatban láthatóak. A különböző módosításokért eltérő enzimatikus funkciójú fehérjék felelősek [43]. Bizonyos PTM-ek esetén az enzimek csupán egyetlen csoportot helyeznek az aminosavakra, míg más esetben ugyanazon az aminosavon több módosító csoport kapcsolása is megtörténhet (például az ubikvitináció esetében) [44]. A módosítások önmagukban, de akár kombinációban is értelmezhetőek [43]. Utóbbira a legismertebb példa a hisztonok módosítása a transzkripció elősegítése céljából, ahol a hiszton fehérjék módosításával (lizin acetiláció) befolyásolják a fehérje töltését, ami a DNS letekeredését eredményezi, míg a foszforiláció, acetiláció és metiláció eredményeként a DNS átírásához szükséges fehérjék toborzása történik. A két folyamat együttesen eredményezi a hatékony átíródást [41, 43].

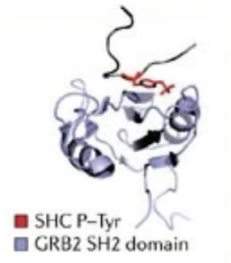
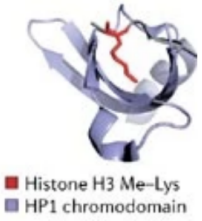
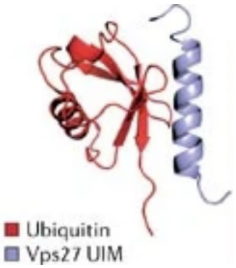
2.3 Fehérje-fehérje kölcsönhatások

2.3.1 Fehérje kötés molekuláris háttere, komplexképződés

2.3.1.1 A fehérje kölcsönhatások kialakulásának feltételei

A fehérje interakciók legalább két fehérje fizikai kontaktusba lépését jelentik molekuláris dokkolás révén, a létrejövő kapcsolat sejtekben vagy élő organizmusokban *in vivo* fordul elő [45]. Az interakció révén kialakuló fizikai kölcsönhatás specifikus, valamilyen funkció ellátására evolválódik [45]. A sejteket sokszor egy, néhány organellumot tartalmazó üres térként képzeljük el, azonban a valóságban maga a sejt molekulárisan zsúfolt [46]. Ebben a hatalmas sűrűségben kell a fehérjéknek megtalálniuk egymást a kölcsönhatások létrejöttéhez. 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. A tömeghatás törvénye kimondja, hogy a fehérje mennyiségét a komplexben a koncentráció és az affinitás együttesen fogják meghatározni. Példaként, ha [A][B] és [A][C] komplex is megvalósulhat, hiába kötődne az előbbi komplex nagyobb affinitással, ha [C] sokkal nagyobb mennyiségben van jelen normál sejtes körülmények között, akkor az [A][B] komplex nem, vagy csak nagyon kis mennyiségben fog létrejönni. Az egyes fehérje komplexek kialakulásnak dinamizmusa tehát függ az azt alkotó fehérjék expressziójától (milyen mennyiségben termelődik), a degradációjától (milyen mennyiségben bomlik le) és a lokalizációjától (azaz hogy ugyanabban a térrészben vannak-e jelen). A kölcsönhatások megvalósulását azonban tovább árnyalhatják egyéb szabályozási mechanizmusok,

2. táblázat A leggyakoribb poszttranszlációs módosítások[41, 44]

PTM	Foszforiláció	Metiláció	Acetiláció	Ubikvitináció
Példa funkció	Szignáltranszdukció, enzim reakció, fehérje-fehérje, fehérje-ligand interakció	Szabályozza a fehérje aktivitást, fehérje-fehérje és fehérje-nukleinsav kölcsönhatásokat, kromatin dinamikát és génaktivitást (hiszton módosítás)	A fehérje lokalizációja és aktivitása, részt vesz a fehérje-fehérje és a fehérje-membrán kölcsönhatásokban	Fehérje degradációs szignál, részt vesz a fehérje-fehérje kölcsönhatásokban
Szerkezeti példa				

DOI:10.15774/PPKE.ITK.2023.001

DOI:10.15774/PPKE.ITK.2023.001

például a korábban említett poszttranszlációs módosítások (ld. 2.2.5. fejezet), az alternatív splicing vagy a kofaktorok jelenléte.

2.3.1.2 A fehérje kölcsönhatások molekuláris alapja

A fehérje-fehérje kölcsönhatásokat ugyanazok az erők stabilizálják, amelyek a monomer fehérjék feltekeredésében is szerepet játszanak. Ezek a viszonylag gyenge kölcsönhatások fogják összetartani úgy a szerkezeteket, hogy a lehető legtöbb H-híd kötés valósuljon meg, és hogy a felszín a legkevésbé nyújtson teret az oldószerek számára. Bizonyos kölcsönhatásban a résztvevő felszínek molekuláris leírásának egyik lehetséges módja az O-gyűrű elmélet: mutációs vizsgálatok alapján megállapították, hogy léteznek 'hot-spot' aminosavak, illetve energetikailag kevésbé fontos aminosavak amelyek körbeveszik őket [46]. A kötésekben résztvevő aminosavak különböző modulokat (klasztereket) alkotnak [46, 47], és a klasztereken belül számos interakció jön létre, míg a klaszterek közötti interakcióknak sokkal kisebb a valószínűsége. Ennek eredményeként egy teljes klaszter törlése a felületről a vártnál kisebb szerkezeti és energetikai hatást fejthet ki, a klasztert alkotó egyedi mutációk additív értékeihez képest. Mivel a hotspot reziduumok evolúciósan konzerváltak, az azonos régióhoz több kötő partnert megengedő fehérjék esetében a fehérjék hasonló klaszter kombinációkat fognak használni [46, 47]. A fehérje-fehérje kölcsönhatások kialakulásának termodinamikai alapja az, hogy a létrejövő komplex alacsonyabb szabad energiával rendelkezik, mint a nem kötött formában lévő fehérje [46].

2.3.1.3 Fehérjék natív szerkezetének és interakcióik összefüggése

Az interakciók kialakításában többféle fehérjeszakasz is részt vehet: domének és rövidebb-hosszabb rendezetlen szakaszok. Ezen régiókon belül a tényleges atomi kontaktust kialakító részek a fehérje felszínén találhatóak. A különböző interakciós lehetőségek közül a legtöbbet a domének révén kialakított kötésekről tudunk. Az interakciós domének (ld. 2.2.2. fejezet) tipikusan 35-150 aminosav hosszúságúak. A jelenleg ismert domének összesen több százezer bináris interakciót és ∼10 000 komplexet képesek létrehozni [48, 49]. A doménekkel szemben a rendezetlen régiók funkcionális előnye, hogy a szerkezeti flexibilitás miatt képesek több, extrém esetekben akár több száz partnerhez is kötődni [50]. A rövidebb rendezetlen szakaszok, lineáris motívumok is számos interakció kialakításáért felelősek, azonban ezen kapcsolatok általánosságban jóval kevésbé feltártak. Jelenlegi elképzelésünk szerint akár egymillió ilyen interakció is előfordulhat eukarióta sejtekben, de ezek töredékét igazolták csak kísérletesen [48]. A különböző rendezetlen részek és a domének egymással különféle kombinációkban (domén-domén, domén-rendezetlen, akár rendezetlen-rendezetlen [51]) képesek kölcsönhatásokat létrehozni. A natív szerkezet a kötés során képes megváltozni - domének esetén a változás kisebb, míg rendezetlen fehérjék esetében a flexibilis rész gyakran felvesz egy stabil szerkezetet. Ezen kívül összetett, pl. több doménes fehérjék esetében előfordulhatnak nagyobb léptékű, a szekvenciában távol eső részeket érintő szerkezeti átrendeződések is.

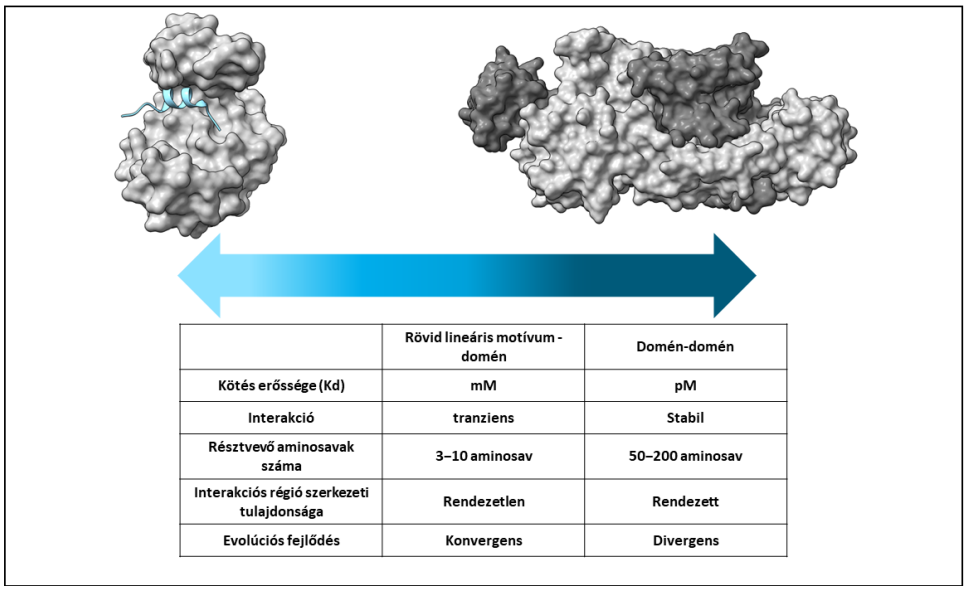
DOI:10.15774/PPKE.ITK.2023.001

2.3.1.4 A fehérje kölcsönhatások típusai a molekuláris összetétel, affinitás és komplexek életidejének függvényében

A fehérjéknek a víz által is hozzáférhető felszíne az a rész, ahol képesek egymással közvetlen fizikai kapcsolatokon keresztül kötni. Két fehérje kapcsolódását bináris kölcsönhatásnak nevezzük, az ilyen kölcsönhatás eredménye egy dimer. A fehérje kölcsönhatásokat osztályozhatjuk a résztvevő fehérjék alapján is, ahol két azonos fehérje esetén homodimer komplexről, míg két eltérő esetén heterodimerről beszélünk. A kapcsolatban résztvevő fehérjék száma lehet magasabb is (trimer, tetramer, pentamer stb., általánosságban: oligomer), ahol a kölcsönhatás megvalósulhat egy időben, vagy az alacsonyabb oligomerizációjú komplexek összeszerelődésével is. Ezekben a komplexekben a fehérjék direkt és indirekt fizikai kapcsolatban állnak [52]. A direkt és indirekt kölcsönhatások meghatározására eltérő kísérletes módszerek léteznek (ezeknek kísérletes meghatározását ld. 2.3.3. fejezet). Az élő szervezetekben a fehérjék legnagyobb része oligomer formában van jelen. Ezt a szerveződési szintet nevezhetjük a fehérjék negyedleges szerkezetének is (ld. 2.2.2. fejezet). *E. coli* esetében például az oligomer formában előforduló szerkezeteket 75%-ra becsülik [15]. Több hipotézis is létezik arra, hogy a fehérjék jelentős része miért oligomer formában fordul elő, az egyik valószínű magyarázat, hogy a nagyobb komplexek ellenállóbbak a denaturációval szemben és a monomer fehérjékkel szemben kisebb hidrofób felszínük van [46]. A fehérje-fehérje interakciók eltérő affinitással alakulnak ki (4. ábra). 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 kJ/mol tartományba esik [46]. 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 valósulnak meg. Általánosságban a komplexek képződése többféle mechanizmus is szerint is létrejöhet.

1. Az egyik a kulcs-zár (lock-and-key) modell, ahol már előre "előkészített" kötési felszínek találkoznak egymással. Ez a mechanizmus a kísérletesen megoldott szerkezetek felénél figyelhető meg és elsősorban fehérje-ligand kötések kialakulását jellemzi. Alapvetően merev kötő partnereket feltételez, de jelen tudásunk szerint bizonyos fokú dinamika - már csak a fehérjék általános mozgékonysága miatt is - szerepet játszik ilyenkor is.
2. Egy másik mechanizmus az indukált illeszkedés (induced fit) modell, ahol a specifikus kötés létrejöttéhez a fehérjének bizonyos konformációs változásokon kell keresztül mennie a kötődés hatására.
3. A harmadik modell a fluktuációs fit (fluctuation fit) néven is ismert konformációs szelekció, ezen elmélet szerint a reakcióba lépésre előkészülő fehérje több konformációs állapotban vár és a legjobban illeszkedő lesz majd képes tényleges interakciót kialakítani [46].

DOI:10.15774/PPKE.ITK.2023.001



	Rövid lineáris motívum - domén	Domén-domén
Kötés erőssége (Kd)	mM	pM
Interakció	tranziens	Stabil
Résztvevő aminosavak száma	3–10 aminosav	50–200 aminosav
Interakciós régió szerkezeti tulajdonsága	Rendezetlen	Rendezett
Evolúciós fejlődés	Konvergens	Divergens

4. ábra A fehérje interakciók típusai a kötések erőssége alapján A skála két végén a tranziens lineáris motívumokkal megvalósuló gyenge és rövid ideig (millisec) tartó interakciók, míg a másik végén a stabil domén domén interakciók hosszú felezési idővel (akár napok).

2.3.2 A fehérje-fehérje interakciók a sejt működésének alapjai: posztszinaptikus példák

A fehérjék kölcsönhatása alapvető fontosságú az élő rendszerekben végbemenő összes folyamatban. A fehérjék legtöbbször nem monomer állapotban vannak jelen, hanem különféle kapcsolatokat alakítanak ki más molekulákkal, sokszor másik fehérjékkel is. Az így létrejött kölcsönhatások révén töltik be funkciójukat, azaz, hogy enzimként, transzporterként vagy akár a sejt stabilizátorokként működjenek [53, 54]. A fehérjék tényleges funkcionalitása azonban nem csak egy-egy komplex működésének eredményeként jön létre. A tágabb molekuláris kontextust értelmezve a fehérjék bonyolult, egymással sokszor összefüggő és együttműködő fehérje-fehérje interakciós hálózatokat alkotnak. Az elemi fehérje a komplex- (és hálózat)szintű jelenségek révén felelős az összetett biológiai funkciókért. Emellett fontos megemlíteni, hogy a fehérjék mellett számos más molekuláknak (pl. RNS-nek, lipideknek) is fontos szerepe van. A következőben a fehérje interakciókon keresztül létrejövő funkciók közül mutatok be néhány posztszinaptikus példát. Ezekben az esetekben az kölcsönhatásokon keresztül egész fehérje gépezetek jönnek létre a feladatok ellátására.

(A) Funkció a fázisszeparációban

A fázisszeparáció viszonylag újonnan felfedezett jelenség, amelynek kiemelt szerepe lehet a posztszinapszisban. A folyadék-folyadék fázisszeparáció (liquid-liquid phase separation, LLPS) lényege, hogy gyorsan változó, membrán nélküli szubcelluláris kompartmentek alakulnak ki, amelyekben lokálisan bizonyos speciális fehérjék feldúsulnak. Ma már több fehérjét is ismerünk, amelyek részt vesznek ebben a folyamatban, többek között a két legnagyobb mennyiségben előforduló posztszinaptikus állványfehérjét,

DOI:10.15774/PPKE.ITK.2023.001

a SynGAP-ot és a PSD-95-öt. A koncentrációfüggő fázistranzíció során a SynGAP fehérje homotrimer coiled-coil-t képez, amelyhez több PSD-95 fehérje képes kapcsolódni. Ezt az kölcsönhatás implikálhatja, hogyan alakulnak ki a nano méretű domének a PSD-ben, vagy arra is választ adhat, hogyan képesek egymástól elkülönülni az egy dendrit részen megtalálható serkentő és gátló szinapszisokat kialakító fehérjék (5. ábra, panel A)[55].

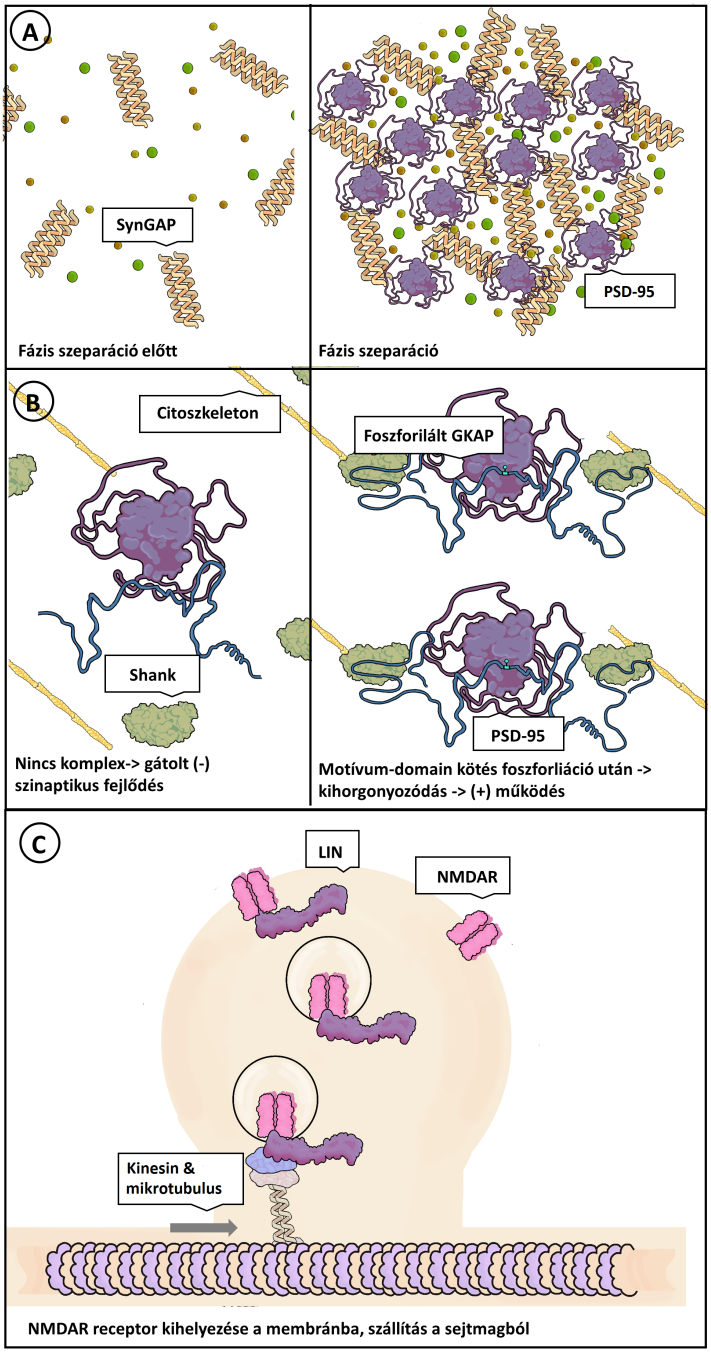
(B) Adapter vagy scaffold funkció

A 1.2. fejezetben már említésre került, hogy a posztszinaptikus denzitásban egy rendkívül fontos fehérje osztályt képeznek a scaffold fehérjék. Az egyik legkiemelkedőbb és legfontosabb ilyen fehérje a PSD-95. A PSD-95 az AMPA sejtfelszíni receptorok kötésért felelős. Ez a receptor veszi fel a preszinaptikus jeleket, és az első lépést jelenti az információ továbbításában a posztszinapszisban. A PSD-95 a a PSD mélyebb rétegében is rendelkezik interakciós partnerrel, például a GKAP (SAPAP1) fehérjével. A PSD-95-ben található C terminális GK domén a SAPAP-ok erősen konzervált RxxSYxxA motívumával képes kölcsönhatásba lépni. Ez az interakció azonban csak abban az esetben valósulhat meg, ha a SAPAP fehérjék foszforilálódnak, ugyanis a kölcsönhatás a PSD-95 GK doménjének kanonikus foszfopeptid-kötő zsebén keresztül valósul meg. A SAPAP fehérjék a PSD-ben a Shank fehérjékkel is kapcsolatban állnak, amelyek aztán az aktin citoszkeletonhoz horgonyozzák a komplexet. Abban az esetben, ha a foszforiláció nem valósul meg, a hármas komplex (PSD-95-SAPAP-Shank) nem tud összeállni. Ezen kölcsönhatások hiányának eredményeként a dendritek nem megfelelően stimuláció hatására fognak kialakulni, ami gátolja a szinaptikus fejlődést (5. ábra, panel B)[56].

(C) A fehérjék lokalizációjában és szállításában betöltött funkció

Ahogy a sejtek komplexitása egyre nőtt az evolúciós fejlődés során, és ahogyan megindult a kompartmentalizáció, egyre fontosabbá vált a különböző molekulák megfelelő szubcelluláris lokalizációba való szállítása [49]. Ezek a folyamatok a már tárgyalt fehérje-fehérje kölcsönhatásokon kívül fehérje-lipid interakciókon keresztül valósulnak meg [49]. A serkentő szinapszisokban megtalálható egyik fontos receptor az NMDA, amely heteromer összeszerelődéssel jön létre az NR1, NR2 illetve NR3 alegységekből [57]. Ezek a neuronok endoplazmatikus retikulumában keletkeznek, ami pedig a neuronok sejtestében található. Innen kell eljutniuk egészen a dendritek végéig [58]. A receptorok szállítását egy molekuláris gépezet fogja végezni, amiben részt vesznek (i) adapter fehérjék (pl. LIN); (ii) szállításban résztvevő fehérjék: (a) Kinesin motorfehérje (b) citoszkeletális hálózatot alkotó mikrotubulus fehérjék; (iii) scaffold fehérjék pl. SAP102. A szinapszisok kialakulása során a NMDAR-t tartalmazó kargó, amelyet az adapter fehérjék a motor fehérjéhez kötnek a citoszkeletális mikrotubulus rendszeren keresztül jutnak majd a dendrit tüskékhez. A végleges PSD lokalizáció elnyeréséért a scaffold fehérjék felelnek. A megnőtt neuronális aktivitás hatására több NMDAR fog kihelyeződni a membránba az említett folyamat révén, aminek eredményeként erősödik a kapcsolat a neuronok között. Ez a folyamat a memória kialakulásának elemi része (5. ábra, panel C)[57].

DOI:10.15774/PPKE.ITK.2023.001



5. ábra A fehérjék interakcióival összefüggő funkciói három posztszinaptikus példán A: Funkció a fázis szeparációban; B: Adapter vagy scaffold funkció; C: A fehérjék lokalizációjában és szállításában betöltött funkció

DOI:10.15774/PPKE.ITK.2023.001

2.3.3 Kísérletes módszerek az interakciók meghatározására

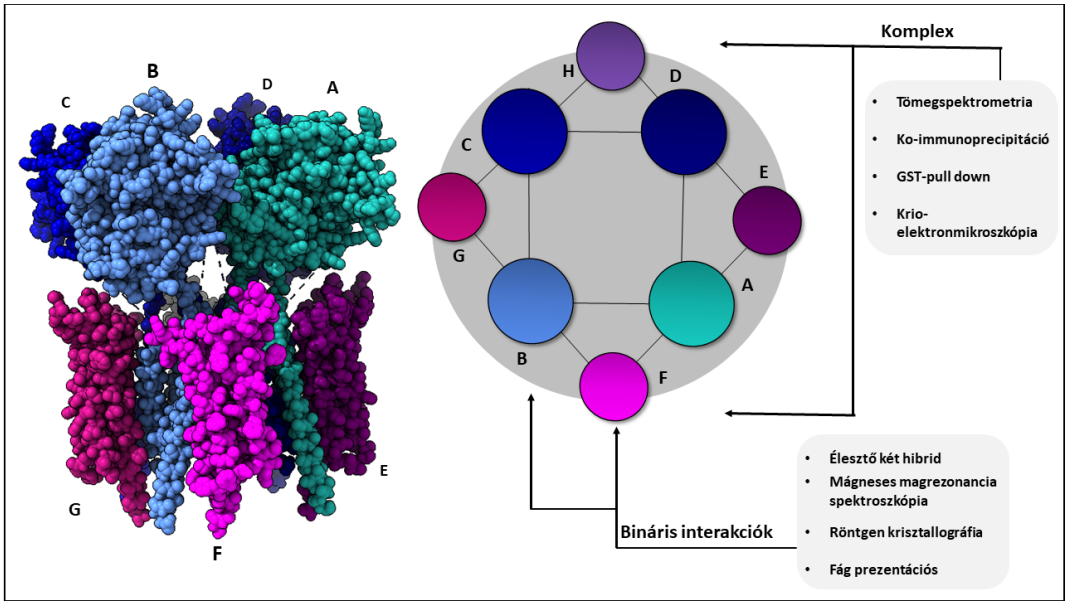
A fehérje-fehérje interakciók kimutatására számos lehetőség van. A különböző technikák eltérő adatokat szolgáltatnak a kölcsönhatásokról, emellett a megbízhatóságuk is eltérő egy-egy hipotézis megerősítésére (3. táblázat). A technológiák jellemzői alapján az alábbi táblázatban láthatóak a legfontosabb kísérletes technikák, amelyek rendelkezésre állnak a fehérje interakciók meghatározására.

3. táblázat A fehérje-fehérje interakciók kimutatására szolgáló technikák, a módszerek típusa szerint csoportosítva

	Direkt interakciók	In vivo	Alacsony affinitású interakciók is	Nagy áteresztőképességű
Biofizikai				
Fluoreszcencia polarizáció (FP)	X			
Felületi plazmon rezonancia (FPR)	X			
Mágneses magrezonancia spektroszkópia (NMR)	X		X	
Röntgenkrisztallográfia (X-ray)	X		X	
Izotermális titrációs kalorimetria (ITC)	X			
Tömegspektrometria (MS)	X			X
Krio- elektronmikroszkópia (Krio-EM)			X	
Biokémiai				
Förster-féle rezonáns energiaátadás (FRET)			X	
Biolumineszencia rezonancia energia transzfer (BRET)			X	
GST-pull down				X
Ko- immunoprecipitáció (coIP)				
Genetikus				
Fág prezentációs	X	X	X	X
Élesztő két hibrid (y2h)		X		X

DOI:10.15774/PPKE.ITK.2023.001

A fenti technikákból kiindulva kombinált eljárások is léteznek [49]. Különböző technikák állnak rendelkezésre bináris fehérje-fehérje interakciók meghatározására, illetve molekuláris komplexek azonosítására [45]. A bináris interakciók meghatározására használható technikák a következők: mágneses magrezonancia spektroszkópia, röntgenkrisztallográfia, élesztő két hibrid, fág display. Ezekben az esetekben a fehérjék közvetlen fizikai kontaktusait (bizonyos esetekben atomi szinten) teszteljük, ill. látjuk. A tömegspektrometria, GTS-pull down és ko-immunoprecipitáció eljárások a komplexek és indirekt kapcsolatok (ld. 2.3.1.4. fejezet) kimutatására használhatóak (6. ábra). Az egyik legfontosabb interakciós adatbázis az IntAct (ld. 2.3.4. fejezet) alapján a legtöbb eredmény az élesztő-két-hibrid, illetve a tömegspektrometriai eljárások valamelyikéből származik. Előbbi fehérje-fehérje interakciók, utóbbi komplexek kimutatására alkalmas (ld. 2.3.1.4. fejezet).



6. ábra Az AMPA receptor- TARP komplex példája a direkt és indirekt interakciókra és a lehetséges kísérleti technikák az interakciók kimutatására (PDB: 5VOV) A szerkezet egy hetero-octamer, amelyben az AMPA receptor alegységek (A,B,C,D láncok) képezik a komplex magját. Hozzájuk asszociálnak a TARP fehérjék (E, F, G, H láncok), amelyek egymással nem lépnek interakcióba

2.3.4 Interakciós adatbázisok és az interakciós adatok rendszerezésének alapjai

A nagy áteresztőképességű technikák elterjedésével és a kísérletesen meghatározott interakciók számának exponenciális növekedésével egyre fontosabbá válik a kapott adatok megfelelő validálása, rendszerezése és eltárolása [45]. A fehérje-fehérje interakciós adatbázisokban általában háromféle adat található meg:

- kísérletesen igazolt interakciók, amelyeket vagy kurátorok visznek be az adatbázisba, vagy közvetlenül a kísérletek megvalósításában résztvevő kutatók küldik be
- számítógépes módszerekkel becsült interakciók,

DOI:10.15774/PPKE.ITK.2023.001

illetve ezek egyesítése. Elsődleges adatbázisnak nevezzük azokat az adatszetteket, amelyek önmaguk állítják össze az interakciós adatokat és meta-adatbázisnak vagy másodlagosnak azokat, amik más adatbázisok adatait integrálják egymásba [59]. Léteznek specifikus adatbázisok is, amik például csak bizonyos fajokból származó adatokat gyűjtenek, vagy amik más tulajdonság alapján (például fehérjefunkció) szűrnek. Az interakciós adatgyűjtés egy viszonylag bonyolult és komplex folyamat. Az irodalomban elszórtan találhatóak meg az interakciós adatok és ezért sokszor algoritmusok segítségével (ún. text mining) gyűjtik össze a releváns publikációkat. A nyilvánvaló előnyök mellett azonban a keresőmotorok tulajdonságai nagyban befolyásolhatják a keresés eredményét [59]. Az adatfeldolgozási folyamat időtartamát és nehézségét az is jelentősen befolyásolja, hogy pontosan mely adatok kerülnek rögzítésre egy-egy publikációból. Ezek a kihívások bizonyos módszerekkel megkönnyíthetőek. Az International Molecular Exchange Consortium-ot (IMEX) azért hozták létre, hogy az interakciós adatokat gyűjtő szervezetek számára kollaborációs lehetőséget biztosítsanak [45]. Az alapvető törekvéseik közé tartozik, hogy redundanciamentes és egyszerűen hozzáférhető adatkészleteket biztosítanak a kutatók számára. Ez több szempontból is lényeges. Egyrészt a kurációs munka bonyolultsága és időtartama miatt érdemes az erőforrásokat hatékonyan elosztani. Egy másik fontos elképzelés közös kurációs szabályok felállítása, amelyhez minden társadatbázis tartja magát. Ez homogén kurációt és jobb minőségű adatokat jelent [60]. Az interakciós adatokat gyűjtő aktív tagok többek között a DIP (https://dip.doe-mbi.ucla.edu/dip/Main.cgi), IntAct, Mint, MatrixDB, I2D és az InnateDB (forrás: http://www.imexconsortium.org/about/) A fentiek mellett az IMEX konzorcium fontos elhatározása, hogy az interakciós adatok tárolását közös formátumban valósítják meg [60]. A “The minimum information required for reporting a molecular interaction experiment” (MIMIx) [61] olyan iránymutatást tartalmazó leírás, ami egy molekuláris interakció leírásának minimális feltételeit szemlélteti. A kinyerni kíván adatok több részre oszthatóak, ahol az elsődleges a vizsgálandó molekulával kapcsolatos információk. Ez a legnagyobb adatvesztéssel járó folyamat, amikor a kísérletet leíró publikációt az adatbázisba gyűjtik. Számos esetben a molekula pontos, egyértelmű azonosítása nem lehetséges, mivel a publikáció nem tartalmazza a gén nevét (és/vagy pontos azonosítóját pl. a UniProt adatbázisból) és az organizmust, amiből a fehérje származik. Kurátorok visszajelzéseit elemezve arra a következtetésre jutottak, hogy a kurációs idő ∼70%-a a megfelelő azonosítók megtalálásával telik. A másodlagos információ a molekula szerepe a kísérletben biológiai értelemben (pl. enzim vagy szubsztrát) és kísérletes értelemben (pl. csali vagy préda), valamint hogy milyen fajban figyelték meg (de lehet *in vitro* is) és hogy milyen kísérletes módszerrel határozták meg az interakciót, beleértve a résztvevő molekulák azonosítását is [61]. A különböző módszereket és leírásokat ontológiákba foglalták, így biztosítva az egységes leírást. Az IMEX konzorcium alapvetően a mély kurációt támogatja [60, 62], amelynek lényege, hogy részletes információt adjon a kölcsönhatásokról, kísérletekről és körülményekről [62] Ezen információk gyakran elengedhetetlenek az interakció kontextusának megértésében és a megfelelő

DOI:10.15774/PPKE.ITK.2023.001

megbízhatósági szint meghatározásában [60]. A MIMIx-en túl extra információt biztosítanak a technikai részletek lehető legrészletesebb leírásával, a kötő régiók, fehérje tag-ek, illetve mutációk rögzítésével. Emellett funkcionális környezet is adnak, például a szubcelluláris lokalizáció megadásával [62]. Összességében az IntAct szintű mély annotáció végső célja többek között az információ olyan mélységű tárolása, hogy ha egy kurátor már dolgozott egy publikáción, azt többet másnak ne kelljen elolvasnia.

2.4 A biológiai funkció megzavarása: mutációk és betegségek

2.4.1 Mutációk és kialakulásuk mechanizmusai

A replikáció az egyik legfontosabb funkció a sejtekben. Ez a folyamat elképesztően precíz, a mutációk gyakorisága 10^{-10} mutáció/bázispár sejtosztódásonként [63]. Azonban nem teljesen hibátlan, és a hibák hatására mutációk alakulhatnak ki. Abban az esetben, ha ezek a hibák csírasejt-vonalban fordulnak elő, akkor azok az utódoknak és azok minden sejtjének is átadódnak [64, 65]. A legtöbb mutáció azonban nem ekkor keletkezik, hanem a zigóta sejtekből kifejlődő sejtek életideje alatt, ezek az ún. szomatikus mutációk. Ebből következően a posztzigótikus mutációk csak a sejtek egy bizonyos populációját érintik, ellentétben a csíravonal mutációkkal, ahol az összes sejt hordozza a variációt [65]. Míg a csíravonal mutációk statikusnak tekinthetők, addig a szomatikus mutációk dinamikusan külső és belső hatásokra megjelenhetnek az egymás utáni replikációk során. A mutációknak több osztálya létezik az alapján, milyen hiba okozza, és hogy milyen következményekkel fog járni. A legegyszerűbb eset az egypontos variáció (single nucleotide variation, SNV). Ezen mutációk gyakorisága 10^{-8} SNV mutáció bázis páronként/generációnként. Emellett léteznek különböző struktúrális, nem egy pontra, hanem hosszabb DNS szakaszokra kiterjedő elemek is: a copy number variation (CNV) 50 bp-nál hosszabb duplikáció vagy deléció, az aneuploidiák (egy teljes kromoszóma megkettőződése vagy deléciója), az indel-ek, amelyek rövid szakaszokra korlátozódnak [64, 66].

Megállapítható 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 [64]. Lesznek tehát kifejezetten előnyös mutációk, amik a spektrum pozitív végén helyezkednek el, ezek szelekciós előnnyel fognak járni. Léteznek olyan mutációk is, 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 csíravonalbeli betegséget okozó mutációk általában egy gyengébb fenotípust eredményeznek, ami negatívan befolyásolja az életminőséget, de nem végzetes. Ennek az az oka, hogy a csíra vonalban megjelenő, komoly funkcióvesztéses mutációk az esetek nagy részében az egyedfejlődés korai szakaszában

DOI:10.15774/PPKE.ITK.2023.001

kiszelektálódnak, így a populáció szintjén nem jelennek meg. A skála negatív végén lesznek azok a (leggyakrabban szomatikus, azaz szerzett) mutációk, amelyek az élettel összeegyeztethetetlen változást okoznak.

2.4.2 Mutációk hatása a fehérje szerkezetekre

Annak függvényében, hogy a mutáció egy pontot érint vagy egy hosszabb szakaszt, eltérő hatással lesz a fehérjére. Utóbbi esetben a változás olyan kiterjedésű DNS szakaszokat érinthet, hogy az egyedi fehérjékre gyakorolt hatás nem értelmezhető. Ezért a következőkben csak az SNV-k hatását kívánom ismertetni. A DNS szintjén megvalósuló mutációk transzláció után különbözőképpen nyilvánulnak meg. A genetikai kód degeneráltsága miatt számos nukleinsav szintű mutáció meg sem jelenik majd az aminosavak szintjén. Azok azonban, amelyek a fehérjék szintjén is megmaradnak, missense mutációkat vagy nonsense mutációkat hozhatnak létre. A nonsense mutáció egy korai stop kodon beépülését jelenti a szekvenciába, amely hatásának kiküszöbölésére már RNS szinten is léteznek folyamatok, amik a hibás átíródás termékeit lebontják, például ilyen folyamat a nonsense-mediated decay [67]. Egyes esetekben azonban így is létrejöhet a fehérje, ami azonban egy rövidebb, csonkolt formában fog szintetizálódni. Ebben az esetben a fehérje nem fogja tudni felvenni a funkciójának betöltéséhez szükséges szerkezet és/vagy az eredeti funkció betöltéséhez szükséges régiók - akár teljes funkcionális domének - hiányozni fognak belőle. Erre a jelenségre az egyik legismertebb betegség a cisztás fibrózis, ahol a CTFR csatorna nem megfelelő konformációja miatt a kloridionok (és a nátrium, illetve víz) nem képesek a membránon keresztük való átjutásra, ami komoly szervi problémákat okoz, a nyák besűrűsödik, mely további fertőzések kialakulásához vezet. A cisztás fibrózisban szenvedő betegek várható élettartama kezelésekkel jelenleg átlagosan 30 és 40 év között várható. [68].

A missense mutációk esetében sokkal megengedőbb a szervezet. A betegséget okozó örökölt missense mutációk (disease-causing germline mutations, DM) többsége a fehérje stabilitását befolyásolja, csupán kis részük befolyásolhatja a kölcsönhatásokat [69]. Stabilitást csökkentő hatás lehet a hidrogénhíd kötések elvesztése [69], vagy az eltemetett régiókban megváltozó aminosavak okozhatnak térbeli ütközéseket [70]. A DM-ek sokszor a negyedleges szerkezetet kialakításáért felelős kölcsönható felszínre esnek [71]. Például a CAMKII fehérjének fontos szerepe van a szinapszisokban, különösen nagy koncentrációban fordul elő neuronális sejtekben és olyan fundamentális jelenségekhez kötötték funkcióját, mint a memória kialakulása. Egy recesszíven öröklődő mutáció a fehérjében, a (His477Tyr) mutáció a CAMKII dimerizációs felszínére esik. Korábbi kutatások már kimutatták, hogy a fehérje dimerizációja elengedhetetlen a fehérje szinaptikus lokalizációjához és a megfelelő szubsztrátok foszforilációjához. A funkció vesztő mutáció a fehérjében közvetlenül felelős a súlyos betegség kialakulásáért [72].

DOI:10.15774/PPKE.ITK.2023.001

2.4.3 Rosszul feltekeredő fehérjékből következő neuronális betegségek

Az előbb említett példa mellett számos ismert és jelentős neuronális betegség hátterében a fehérje misfolding áll (4. táblázat). A feltekeredés hibájának legnagyobb hatása a neurodegeneratív betegségek patomechanizmusa esetében mutatható ki. A nem megfelelően feltekeredett fehérjék aggregátumokat képezve stressz- és neurotoxikus folyamatok révén negatívan befolyásolják a kognitív működést. A fehérje misfolding hátterében örökletes mutációk és egyéb tényezők is állhatnak, például az öregedéssel együtt járó megváltozott celluláris folyamatok. Pontos tudásunk ezen betegségek okairól és molekuláris szintű lefolyásáról azonban - több évtizedes intenzív kutatások ellenére - még nincsen [73, 74]. A jelenlegi orvostudomány egyik nagy kihívása a hibás folding eredményeként megjelenő fehérjék által okozott tünetek csökkentése, esetlegesen a kórkép szinten tartása, visszaszorítása. A farmakoterápiák többsége a korai stádiumban csökkenthetik, sőt, lassíthatják a betegség kialakulását, azonban vissza nem fordítják azokat.

4. táblázat Jelentősebb neurodegenratív betegségek és a legfontosabb/legtöbbet tanulmányozott aggregátumot képző fehérjék, misfolding feltételezett oka [73]

Betegség	Misfolded fehérje	Feltételezett elsődleges ok
Alzheimer	Aβ42	öregedés folyamatához köthető mutációk
Alzheimer	Tau	hiperfoszforiláció
Huntington	mHTT	autoszomális domináns mutáció
Parkinson	α-synuclein	örökletes és sporadikus missense mutációk, oxidatív stressz
Amyotrophic lateral sclerosis	Superoxide dismutase	örökletes mutáció
Spinocerebellar ataxia	több fehérje	autoszomális domináns mutáció

Mára már számos módszert fejlesztettek ki, amelyek a fehérjék szerkezeti adatait és a mutációs információkat a pozitív-semleges-negatív hatás alapján osztályozza [75]. A pontos analízishez azonban szükség van kísérletileg elérhető térszerkezetekre, amely sok fehérje esetében még nem elérhető. A nagy mennyiségű genomszekvenálási adat alapján, illetve a szerkezetbecslő eljárások fejlődésének következménye, hogy a mutációk funkcionális elemzésére egyre több lehetőség adódik, ami a betegségek kezelésében is új távlatokat nyithat meg

DOI:10.15774/PPKE.ITK.2023.001

3 Célkitűzés

Számos olyan alapvető agyi működés és funkció van, amelyet teljes molekuláris részletességgel ma még nem ismerünk. Az ezekben résztvevő jelátviteli folyamatok közül számos a posztszinaptikus oldalon, az itt lokalizálódó fehérjék kölcsönhatásain és funkcióin keresztül valósul meg. A doktori dolgozatom célja a posztszinapszis fehérjéinek részletesebb megismerése elsősorban szerkezeti és funkcionális szempontból a fehérje bioinformatika eszköztárát használva. Ehhez számos előre megfogalmazott célkitűzésünk volt, azonban a kutatás során az eredmények függvényében új irányok is megfogalmazódtak.

1. A posztszinapszis pontos összetétele ma még nem ismert, és nem létezik olyan specifikus adatforrás, amely független adatokat kombinálva tartalmazna megbízható adattárat ezen fehérjék számára. Ezért az első feladat azon fehérjék körének meghatározása volt, amelyek szerepet játszanak a szinaptikus jelátvitel posztszinaptikus oldali jelátvitelében.

2. Mivel az egyéni variációk és a neuronális betegségek kialakulásában fontos szerepe lehet a posztszinapszis fehérjéinek, ezért az ismert és olyan megbízható variánsok és mutációk összegyűjtése is kitűzött cél volt, amely pozíciója egyértelműen meghatározható a fehérjék szekvenciájában és szerkezetében.

3. A posztszinapszisban jellemzően nagyméretű komplexek alakulnak ki és ezek felelősek az eddig feltárt jellegzetes funkciók kialakításáért. Mi a posztszinapszis fehérje interakciós hálózatának vizsgálatát szerettük volna megvalósítani olyan részletességgel, hogy a kölcsönhatásokat kialakító szakaszokat is meghatározzuk.

A 3. pontban megfogalmazott kérdéskör elemzése során egy érdekes jelenséget figyeltünk meg, amelynek nagyon szűkszavú irodalmi leírását találtuk csupán. Ezért a jelenség mélyebb elemzésébe kezdtünk, azonban mivel a humán proteom szintű leírás limitált volt, nem csupán a posztszinaptikus fehérjéken vizsgáltuk azt. Az esetleges posztszinaptikus / központi idegrendszeri eshetőségeket azonban mindvégig kiemelt figyelemmel kezeltük. Az itt megfogalmazott irányok a következők voltak:

1. A betegséget okozó variánsoknak sokszor erőteljes fenotípusos megnyilvánulása van, azonban nem járnak élettel összeegyeztethetetlen következménnyel. Ezek a mutációk jellemzően olyan szerkezeti részeken akkumulálódnak, amelyek szerkezeti szempontból jól strukturáltak. A coiled-coilok esetében azonban ennek a jelenségnek éppen az ellenkezőjét tapasztaltuk, amelyet részletesebben is vizsgálni kívántunk elsősorban a coiled-coil szerkezeti sajátosságai szempontjából.

2. Vizsgálni szerettük volna, hogy látunk-e bármilyen neurális betegséggel való összefüggést a coiled-coil szerkezetben való érintettséggel.

DOI:10.15774/PPKE.ITK.2023.001

4 Adatok és módszerek

4.1 Adatok

4.1.1 Uniprot

A UniProt adatbázis a legnagyobb fehérje szekvenciát tartalmazó adatbázis, amely a szerkezetről és a funkcióról is tartalmaz adatokat. Az információk gyűjtését és rendszerezését a területen jártas kurátorok végzik. Az annotációban elérhető információk közül kiemelendők: fehérjék szubcelluláris lokalizációja, variánsok, izoformák, domének és kölcsönhatások. A UniProton belül található meg a Humsavar, ami a UniProton belüli, önálló adatbázis, és humán missense variációkat és azok funkcionális osztályozását tartalmazza.A legtöbb mutációs adatot a DNS szekvenciális adatokból szerezhetjük be, azonban ezek fehérje szintre vetítése rendkívül nehéz feladat, a Humsavar magas szinten annotált adatai aminosavszintű mutációs adatokat tartalmaznak. Alapvetően két forrásból származnak az adatok, vagy a NCBI dbSNP-ből vagy közvetlenül irodalmi adatok feldolgozásából. Mindkét esetben számos megkötést alkalmaznak az adatok átvételekor. Utóbbi esetében különös jelentősége van a neutrális és a betegséget okozó mutációk elválasztásának, amely egy több pontos rendszer alapján történik [76]. A Humsavar általunk használt verziójában (2019), megtalálható információk a következőek voltak: gén neve, Swiss-Prot azonosító, a variáns egyedi annotációkor hozzáadott azonosítója (FTId), az aminosav cserére vonatkozó információ (részletesebben vad típus aminosava, aminosavpozíció, mutáns aminosav), a variáció típusa (polimorfizmus, betegséget okozó mutáció vagy nem osztályozott), dbSNP azonosító, betegséget okozó mutációk esetén a betegség neve. A UniProt egyszerre tekinthető elsődleges és metaadatbázisnak, például az expressziós vagy filogenetikai adatokat más, ezekre az információkra specializált adatbázisokból veszi át. Fontos megjegyezni azonban, hogy mivel a UniProt egy általános adatbázis bizonyos specifikus adatokat nem a legjobb minőségben tartalmaz [77].

Elérhetőség: `https://www.uniprot.org/`

Humsavar közvetlenül:

`https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/docs/humsavar.txt`

4.1.2 Posztszinaptikus adatbázisok, adatkészletek: SynaptomeDB, SynGO, G2C

A posztszinapszis pontos fehérje összetételét az agy komplexitása és a kísérletes technikák limitációi miatt jelenleg sem nem ismerjük teljesen. Ezért fontosnak tartottuk, hogy több adatbázisból származó információt integráljunk. Az alábbiakban az általam felhasznált, szinaptomra specializált adatbázisokat tekintem át röviden. A SnyaptomeDB a szinaptikus gének integrált adatbázisa, amelyet külső adatbázisokat és irodalmi adatokat egyesítve, majd azokat annotálva hoztak létre. 2200 proteomikai adatot közlő publikációt átnézve véglegesítették az adatbázist 2012-ben, és azóta rendszeres frissítéseket végeznek

DOI:10.15774/PPKE.ITK.2023.001

rajta azóta is [78].

Elérhetőség: `http://metamoodics.org/SynaptomeDB/index.php`

A Genes2Cognition (G2C) egy konzorciális adatbázis, ahol fiziológiai és patofiziológiai kontextusba helyezve találhatóak szinaptikus fehérjék, ami így egyfajta katalógusként szolgálhat a neurobiológusok számára. A prezentált adatok több különböző kísérletes vizsgálat eredményét ötvözik, amelyeket a konzorcium tagjai végeztek agyi műtétek során gyűjtött humán vagy rágcsáló minták tömegspektrometriai elemzésével (amelyekhez a humán ortológokat társították később) [79].

Elérhetőség: `https://genes2cognition.org/`

A SynGO konzorcium célja a meglévő GeneOntology (ld. 4.3.4. fejezet) leírásokat specifikusabbá tenni a szinaptikus fehérjék számára. Szakirodalmi adatokat felhasználva fejlesztenek megfelelő leírásokat a fehérjék szinapszison belüli lokalizációjához és a szinaptikus funkciójához. Részletessége miatt talán ez az adatbázis a legértékesebb, azonban a többi adatbázissal összehasonlítva ez tartalmazza a legkevesebb adatot [8]. Elérhetőség: `https://syngoportal.org/`

4.1.3 PFAM (Protein Families)

A domének a fehérjék önálló funkcióval is rendelkező építőelemei, különböző kombinációik biztosítják a fehérjék változatos funkcióit (ld. 2.2.2. fejezet). A Pfam adatbázis a fehérjecsaládok és domének leírására törekszik többszörös szekvenciaillesztések és rejtett Markov modellek segítségével. Az illesztések a HMMER eljárás [80] segítségével készülnek, míg a domének azonosítása a CATH [81] és a SCOP [82] hierarchikus szerkezeti adatbázisok alapján történik. A megoldott szerkezetek számának hirtelen növekedése miatt a Pfam a jövőben mély gépi tanulásos modellekre készül átállni [83]. Elérhetőség: `https://pfam.xfam.org/`

4.1.4 PDB (Protein Data Bank)

A PDB a makromolekulák háromdimenziós szerkezeti adatainak globális archívuma, amelyben megtalálhatóak a kísérletesen megoldott térszerkezetek. Három tükör adatbázist hoztak létre ezen adatok elhelyezésére - az ePDB-t (európai), az RCSB PDB-t (amerikai) és a PDBj-t (japán) - melyek információtartalma teljesen megegyezik, azonban a kiegészítő szolgáltatások eltérnek a különböző oldalakon. A PDB az adattárolás mellett rendkívül fontos szerepet játszik a szerkezeti adatok standard gyűjtésében is. A PDB fájlformátumban az egyes atomok koordinátái mellett meghatározott struktúrában a kísérlet és a fehérje szempontjából releváns adatok is megtalálhatóak [84]. Az adatbázis jelenleg fokozatosan áll át a rugalmasabb mmCIF formátum használatára.

Elérhetőségek: `https://www.ebi.ac.uk/pdbe/` (Európa), `https://www.rcsb.org/` (USA) `https://pdbj.org/` (Japán)

DOI:10.15774/PPKE.ITK.2023.001

4.1.5 Fehérje-fehérje kölcsönhatás adatbázisok: BioGRID, IntAct, STRING

A fehérje kölcsönhatás adatbázisok csoportosításainak alapjairól részletesen a 3.4. fejezetben írtam. A BioGRID az egyik legnépszerűbb kölcsönhatási adatbázis, amelyben jelenleg több mint 2 millió interakciós adat található. A bináris interakciókhoz ezen felül evidencia szinteket rendel, amelyet a megbízható kísérletek száma alapján határoz meg. Az interakciókról eltárolt információk: a résztvevő fehérjék, forrás organizmus, a fehérjék kísérletben betöltött szerepe (pl. csali vagy préda), kísérlet és annak besorolása. Az interakciós adatok mellett poszttranszlációs módosításokat is tárol [85]. Elérhetőség: `https://thebiogrid.org/`

Az IntAct az egyik legjelentősebb elsődleges interakciós adatbázis, ami a kölcsönhatások leírásában sok szempontból 'gold standard'-nak tekinthető. A céljuk a részletes és megbízható adatfeldolgozás és -tárolás ('deep curation') támogatása, aminek lényege, hogy a kísérlet mellett minden olyan körülmény is rögzítésre kerül, ami befolyásolhatja a vizsgálat eredményét. Összefoglalva: az IntAct esetében olyan részletességű leírásra törekszenek, ami lehetővé teszi, hogy az IntAct szinten annotált cikkeket többé "ne kelljen újra olvasnia" egy felhasználónak, mert minden információ kinyerhető a bevitt adatokból. Az adatok hierarchikus felépítésűek és ontológiák által szabályozottak. Ez az egyetlen nagyobb adatbázis, ami az interakciós adatokat a kötő régiók feltüntetésével is képes reprezentálni. A kötő régiók osztályozása:

- Szükséges kötőrégió: amire abszolút szükség van (általában mutagenezissel határozzák meg)
- Elégséges kötő régió: a kötés bekövetkezik, ha ez a régió jelen van
- Kötéssel összefüggő régió
- Direkt kötőrégió: a két molekula fizikai érintkezése [62].

Elérhetőség:`https://www.ebi.ac.uk/intact/home`

A STRING integrált adatbázis, aminek célja az összes ismert és becsült fehérje asszociáció összegyűjtése. A legtöbb adat bináris interakciókból származik, amelyet egy megbízhatósági értékkel is ellátnak. Az interakciós adatok mellett koexpressziós információkat is magába foglal, valamint a fehérjék működésének megértése szempontjából lényeges adatokat is tartalmaz: pl. jelátviteli útvonalra vonatkozó és funkcionális adatokat. Mivel számos interakciót ko-expressziós adatokra és homológiára alapozva számolnak, tartalmaz kevésbé megbízható adatokat is, azonban minden interakcióra megbízhatósági értéket számolnak, ami elérhető [86]. Elérhetőség: `https://string-db.org/`

4.1.6 További adatbázisok

4.1.6.1 ELM (Eukaryotic Linear Motif database) Az ELM adatbázis eukarióta lineáris motívumok legmegbízhatóbb forrása, amelyet a terület szakértői annotálnak kísérletes irodalmi adatokat felhasználva. A motívumokat funkciójuk alapján több csoportba sorolják: hasító, degradációs, dokkoló,

DOI:10.15774/PPKE.ITK.2023.001

ligandkötő, poszttranszlációs módosítások által érintett, illetve a fehérje sejten belüli eloszlását irányító motívumok. Az adatbázis mellett az ELM egy predikciós lehetőséget is nyújt, tetszőleges szekvencia (vagy azonosító), szubcelluláris lokalizáció és taxon megadásával az adatbázisban található motívumok becsült előfordulását határozhatjuk meg. A motívum keresés kontextuális szűrőkkel van kiegészítve, és a találatok valószínűségi érték alapján is rendezhetőek [87].
Elérhetőség: `http://elm.eu.org/`

4.1.6.2 PhaSePro Az adatbázis fázisszeparációban résztvevő fehérjéket és régiókat gyűjt irodalmi adatok alapján. Kifejezetten a fázisszeparációban meghatározó, ún. 'driver' szerepet betöltő fehérjéket listázza, emiatt nem feltétlenül tartalmaz minden, a létrejövő asszociátumokban megtalálható molekulát. Információt tartalmaz a fázisszeparációt meghatározó fehérjerégióról is, amennyiben az ismert [88].
Elérhetőség: `https://phasepro.elte.hu/`

4.1.6.3 HTP (Human Transmembrane Proteome) A transzmembrán fehérjék polipeptid lánca egyszer vagy többször áthalad a kettős lipidrétegen. Mivel ezeknek a fehérjéknek a kísérletes meghatározása rendkívül nehézkes, sokszor egy alacsonyabb szintű szerkezeti definícióval szokták őket jellemezni - ez a topológia, amely megadja a transzmembrán szegmensek számát és helyét a szekvencián belül, illetve a köztes hurkok orientációját a membránhoz viszonyítva (külső/belső). A HTP a humán alfa-helikális transzmembrán fehérjék topológia adatbázisa [89]. Elérhetőség: `http://htp.enzim.hu/`

4.1.6.4 OMA Az OMA módszer és adatbázis a teljes genomok közötti ortológok megtalálására és tárolására. Ma már körülbelül 2 500 teljes genomot tartalmaz. Elérhetőek benne páronkénti ortológ fehérjék, illetve hierarchikus ortológ csoportokat is létrehoznak, amelyekben a közös ősi génből származó, ugyanakkor különböző taxonba tartozó fehérjék találhatóak [90].
Elérhetőség: `https://omabrowser.org/oma/home/`

4.2 Felállított adatszettek

4.2.1 PS_STRICT adatszett létrehozása

A posztszinapszis pontos összetétele ma még nem pontosan ismert az agy területenkénti diverzitásának és a posztszinaptikus proteom meghatározásának kísérletes nehézségei miatt. Az egyedi források sokszor zajjal terheltek, viszont ebben az esetben egy megbízható adathalmaz létrehozása volt a cél, amin az egyes statisztikai jellemzők jobban tanulmányozhatóak. A többi adatszetthez (ld. 4.2.4. fejezet) képest szigorú halmaz létrehozásához a SynaptomeDB és a G2C által listázott gének metszetét használtam, ami végül 1113 fehérjét tartalmazott (a letöltés időpontja 2018).

DOI:10.15774/PPKE.ITK.2023.001

4.2.2 Mutációs adatszettek

A Humsavar adatbázisból letöltött mutációs adatokat visszatérképeztem a PS fehérjékre (PS_STRICT) és a humán proteomra (2018). A teljes proteomon belül 25 751 DM-et és 38 689 PM-et lehetett azonosítani, míg a PS-ben 2 181 DM-et és 1 881 PM-et.

4.2.3 Coiled-coil adathalmazok létrehozása ('CC_STRUCT' és 'CC_SEQ' adatszettek)

A coiled-coil régiók meghatározására két adatkészletet hoztam létre. Minden esetben a humán proteomból indultam ki, amin redundancia szűrést végeztem a CD-HIT programmal, hogy a statisztikákat ne befolyásolják a népesebb fehérjecsaládok - így a fizikiokémiai tulajdonságok sokkal jobban vizsgálhatóak. Az első halmaz létrehozásához a PDB adatbázisban található humán fehérjéken lefuttattam a SOCKET programot, amely a szerkezeti tulajdonságok alapján képes meghatározni a coiled-coil régiókat, az egyes aminosavakhoz regisztereket rendel (ld. 2.2.1. fejezet), valamint meghatározza a szerkezeti elem oligomerizációs állapotát ('CC_STRUCT' adatszett). Mivel a PDB-ben nem található meg az összes emberi fehérje térszerkezete, egy második halmazt is létrehoztam, amelynek a pontossága várhatóan elmarad a szerkezeti halmaztól. Itt szekvencia alapú predikciókat használtam (ld. 4.3. fejezet), hogy meghatározzam a szerkezeti adatkészletben is definiált tulajdonságokat. A régiókat a DeepCoil, a Marcoil, az Ncoils és a Paircoil programokkal határoztam meg, azok alapbeállításaival. A DeepCoil esetén a PSSM (position-specific scoring matrix, azaz pozíció specifikus pontmátrix) módot használtam, ahol az egyes fehérjékhez tartozó mátrixot a SwissProt adatbázis alapján hoztam létre, PSI-BLAST használatával, 10^{-5} határértékkel, 3 iterációval. A heptád pozíciók meghatározásához a MarCoil, Ncoils és Paircoil programokat használtam. Az oligomerizációs állapotot a Logicoillal becsültem. Az adatokból kiszűrtem a magányos alfa-helikális régiókat a CSAH server 2.0 [91, 92] segítségével, mely a SCAN4CSAH és az FT_CHARGE eljárások konszenzusát használja ('C_CSEQ' adatszett). Az adatokat 2020 áprilisában töltöttem le a forrás adatbázisokból.

4.2.4 A teljes posztszinapszis meghatározása ('PS adatszett')

A későbbi vizsgálatokhoz szükség volt egy olyan PS fehérjéket tartalmazó adatszettre, amely sokkal engedékenyebb a 4.2.1-ben bemutatott adatoknál. Ennek létrehozásához a már ott is használt két forráson (G2C és SynaptomeDB) kívül a SynGO adathalmazt, valamint a GO Cellular compartment ontológiáját használtam - mindkét esetben azokat a géneket vettem figyelembe, amelyeknél az annotációban szerepelt a 'postsynaptic' kulcsszó. Ez az adathalmaz sokkal engedékenyebb, így az összes olyan fehérjét tartalmazza, amely a 4 forrás bármelyikében szerepel. Az egérben és patkányban található géneket ortológiatérképezéssel visszavetítettem a humán megfelelőre az OMA adatbázis, valamint a génnevek segítségével. Az adatokat 2021 júliusában töltöttem le a forrásadatbázisokból.
Megjegyzés: mivel a különböző adatszettek létrehozása között akár 3-4 év is eltelt, a forrás adatbázisok

DOI:10.15774/PPKE.ITK.2023.001

változásai miatt eltérhet a fehérjék és mutációk pontos listája. Ez az eltérés azonban minden esetben 10%-nál alacsonyabb. Az adatokat hozzáférhetővé tettem az alábbi linkeken

A posztszinaptikus vizsglatok adatai:

`https://github.com/zsofii/dkzs_dis/tree/main/ps`

A coiled-coil vizsgálatok adatai:

`https://github.com/zsofii/dkzs_dis/tree/main/CC`

4.3 Bioinformatikai módszerek, algoritmusok

4.3.1 BLAST (Basic Local Alignment Search Tool)

Az egyik leggyakrabban használt bioinformatikai algoritmus, aminek segítségével a keresett fehérjével / fehérjeszakasszal homológ fehérjéket vagy fehérjeszakaszokat lehet találni egy háttér adatbázisból. Heurisztikus eljárás, az első lépésben a valószínűsíthetően hasonló szekvenciák azonosításához egyszerűsítéseket alkalmaz két- három aminosavas/nukletotidos szakaszok (tuples) előfordulási valószínűsége alapján, melyeket a használt háttér adatbázisokhoz egy előzetesen indexelt fájlban tárol. A második lépésben hagyományos lokális szekvenciaillesztést végez (Smith-Waterman algoritmus). Jelenlegi megvalósításai számos további megoldást alkalmaznak a nagy adatbázisokon való keresés meggyorsítása érdekében, illetve léteznek széles körben használt, az érzékenységet iteratív keresésekkel növelő változatai is (pl. PSI-BLAST). Online és lokálisan futtatható formában is elérhető [93].

Elérhetőség: `https://ncbi.nlm.nih.gov/blast`

4.3.2 Clustal(Omega)

Homológ fehérjék többszörös szekvenciaillesztésének létrehozását teszi lehetővé . A kétlépéses algoritmus először egy 'mindenki mindenkivel' (all-against-all) illesztés segítségével meghatározza az egyes szekvenciák páronkénti távolságát, majd ezt felhasználva iteratív módon illeszti az egyes szekvenciákat a meglévő illesztéshez. Több szekvencia esetében azokat profilként kezelve végzi az illesztést. Kiválóan paraméterezhető, könnyen használható eljárás, ugyanakkor távoli hasonlóságok esetében nem feltétlenül éri el a HMM vagy PSSM-alapú eljárások megbízhatóságát [94].

4.3.3 CD-HIT

Nem redundáns fehérje adatkészletek létrehozására szolgáló algoritmus, ami egy szekvencia gyűjteményből a leginkább reprezentatívnak tekinthető szekvenciákat adja vissza egy, a felhasználó által megadható azonossági küszöb figyelembevételével. Különböző hosszúságú, küszöb feletti hasonlósággal rendelkező szekvenciák esetében a hosszabb megtartását preferálja. Segítségével a felülreprezentált szekvenciáktól mentes, csökkentett méretű adat szetteket kaphatunk. Kifejezetten gyors, hatékony eljárás [95].

DOI:10.15774/PPKE.ITK.2023.001

4.3.4 A GO (GeneOntology) funkcionális osztályozás

A GeneOntology a fehérjék funkcionális elemzéséhez használt egyik legelterjedtebb rendszer. A kezdeményezés célja egy egységes leírás megteremtése gének és géntermékek megfelelő reprezentációjára (bármilyen élő faj esetén). Ehhez megfelelő, pontosan definiált leíró szókészlet-gyűjteményt (controlled vocabulary) fejlesztenek. Az egyes fehérjékre vonatkozó információkat három szempontból rendszerezi:
a) celluláris lokalizációt (cellular component) pl. synapse (GO:0045202),
b) molekuláris szintű funkciókat, amelyek a fehérjére jellemzőek, sokszor más fehérjékkel kölcsönhatásban (molecular function) pl. kináz kötés (GO:0019900)
c) általános célt vagy biológiai folyamatot, amelynek végrehajtásában a fehérje részt vesz (biological process) pl. génexpresszió regulációja (GO:0010468)
Ezeket az információkat (terminusokat) hierarchikus rendszerben helyezi el, ahol a különböző terminusok kapcsolatai irányított módon vannak összekötve [96].
Elérhetőség: `http://geneontology.org`

4.3.5 Coiled-coil szekvenciális predikciós módszerek: DeepCoil, Ncoils, Paircoil, Marcoil, Logicoil

A coiled-coilok előfordulásának becslése a jellegzetes szerkezeti attribútumok miatt a kanonikus heptád ismétlődéseket tartalmazó szekvenciák esetében viszonylag megbízhatónak tekinthető. A különböző módszerek eltérő koncepciókat alkalmaznak. A DeepCoil [97] az egyik legújabb módszer, ami konvolúciós neurális hálókat felhasználva képes kanonikus és nem kanonikus coiled-coilok predikciójára. A korábbi módszerek klasszikus algoritmusokon alakulnak: a PairCoil [98] páronkénti valószínűséget számol, Marcoil [99] rejtett markov modellt használ, míg az Ncoils [100] egy profil (PSSM) alapú kereséseket alkalmaz. A coiled-coil felismerésén kívül egyéb programok is léteznek, amelyek például a coiled-coil oligomerizációs állapotát képesek megbecsülni, pl. LogiCoil [101].

4.3.6 Coiled-coil szerkezeti annotációs módszer: SOCKET

Az algoritmus az oldalláncok coiled-coil szerkezetre specifikus elrendeződését ismeri fel és ennek alapján meghatározza a coiled-coilok pontos kezdeti és végpontját fehérje szerkezetekben, valamint az egyes aminosavakhoz hozzárendeli a regiszterbeli pozíciókat a heptádokra használt a-h jelölés szerint. A fehérjeszerkezet mellett a futtatáshoz szükség van a szerkezethez tartozó másodlagos szerkezeti elemek detekciójára (DSSP formátumban), a program ez alapján azonosítja az alfa-hélixeket, amelyek között a kölcsönhatásokat elemzi [102].

DOI:10.15774/PPKE.ITK.2023.001

4.3.7 FoldX

A FoldX a fehérjeszerkezetek stabilitását becslő algoritmus, amely a fehérje szabadenergia-változását számoljab izonyos hatásokra. A pontmutációk fehérje szerkezetre gyakorolt hatásának becslése mellett a racionális fehérje tervezésben is használható [103].

4.3.8 Egyéb eszközök (IUPred, DiseaseOntology, Jalview, DSSP, PISA)

Az IUPred az egyik legelterjedtebben használt rendezetlen predikciós módszer, amely nagy pontossággal képes becsülni a rendezetlen régióban található aminosavakat a fehérje szekvenciában [104].
A DiseaseOntology a GeneOntologyhoz hasonlóan azonban betegségek leírására teremt leíró rendszereket [105].
Jalview többszörös szekvencia illesztések létrehozására és elemzésére szolgáló módszer [106].
A DSSP másodlagos szerkezet annotációs eljárás, ami a hidrogén kötések alapján rendeli hozzá a különböző másodlagos struktúrákat a fehérjékben található aminosavakhoz. Emellett az egyes aminosavak oldószer általi hozzáférhetőségét is képes megbecsülni [107].
PISA: A PISA (Protein Interfaces, Surfaces and Assemblies) egy interaktív eszköz makromolekulák szerkezetének elemzéséhez. Számos funkcióval rendelkezik, én a negyedleges szerkezet becslésére használtam a PDB szerkezet alapján [108].

4.4 Statisztikai módszerek

4.4.1 Mutációk feldúsulásának elemzése (DM/PM enrichments)

A betegséget okozó mutációk és polimorfizmusok előfordulását különböző szerkezeti régiókban és azon kívül kontingencia táblázatban rögzítettük (az előfordulás alatt az érintett aminosavak számát értjük):

5. táblázat Kontingencia táblázat a mutációk eloszlásáról

	Betegséget okozó (DM)	Polimorfizmus (PM)	Esethányados (Odds ratio)
Pozitív	x1	x2	x3
Negatív	x4	x5	x6

A DM-ek és PM-ek feldúsulását a következő képelettel számoltuk a táblázat alapján: Képlet1: A feldúsulások számolása

1. Enrichment (DMs) = (x4/(x4+x1))/(x6/(x6+x3))
2. Enrichment (PMs) = (x5/(x5+x2))/(x6+/(x6+x3))

Annak vizsgálatára, hogy milyen összefüggés van két tényező között a hatásuk nagyságát (odds ratio) vizsgáltam az alábbi képlettel:

DOI:10.15774/PPKE.ITK.2023.001

Képlet 2: Odds ratio kiszámítása

1. OR=(x1/x2)/(x4/x5)

4.4.2 Szignifikancia teszt χ^2 próbával és Kolmogorov–Szmirnov-teszttel

A kontingencia táblázat alapján a statisztikai szignifikanciát χ^2 próbával vizsgáltuk. A Kolmogorov-Szmirnov-próbát alkalmaztunk a 5.2.2. fejezetben, ahol a coiled-coilok N-terminális 28 aminosava esetében vizsgáltuk a mutációk szekvenciabeli eloszlásának szignifikanciáját.

4.4.3 P-érték korrekció Bonferroni teszttel

A coiled-coil N-terminális DM halmozódás statisztikai sziginfikancia értékét a Bonferroni teszttel korrigáltuk. A Bonferroni korrekció lényege, hogy az alfa szintjét úgy módosítjuk, hogy kontrolláljuk az I. típusú hiba elkövetésének valószínűségét.

4.4.4 Bootstrap + szórásból számolt szignifikancia (DM/PM AS változás)

Az aminosav cserék által okozott fiziko kémiai változásának (ld. 5.2.3. fejezet) szignifikancia becslése bootstrap módszerrel történt. A bootstrap lényege, hogy a kiindulási mintahalmazunkból bootstrap mintacsoportokat hozunk létre, amelyben véletlenszerűen kiválasztva az adatok 80%-át találjuk. Ezt a folyamatot százszor ismételjük, majd az egyes bootstrap csoportokon átlagot és szórást tudunk számolni. A különböző adatszettek így már összevethetőek egymással: a szignifikancia az átlagok és szórások kiszámításával a 68-95-99.7 szabályt figyelembevéve lett meghatározva. A módszer előnye, hogy viszonylag egyszerűen, kis elemszámú kiindulási mintahalmaznál is használható.

4.5 Vizualizáció

A fehérjeszerkezeti ábrákat a Chimera X 1.1 programmal készítettem [109]. Az ábrákhoz és grafikákhoz a Microsoft Office PowerPointot és, Excelt, a Photoshoppot 23.5 és Python3 csomagokat használtam.

4.6 Egyéb

Shannon-entrópia: az információs technológiában megjelenő fogalom, ami egy véges sok jelből (ABC) álló üzenet információtartalmát képes mérni. A fogalom megfeleltethető a fehérjeszekvenciákra is. A dolgozatomban a többszörös illesztéseknél vizsgáltam az egyes pozíciók információtartalmát, hogy megállapítsam egyes aminosavak milyen mértékben konzerváltak.
Hozzáférhetőség : RSA=ASA/maxASA (RSA: relative surface accessible surface area, relatív oldószer által hozzáférhető felszín, ASA: solvent accessible surface area, oldószer által hozzáférhető felszín, MaxASA: maximal solvent accessible surface area, maximum oldószer által hozzáférhető felszín.) Maximális hozzáférhető felszín értékei az alábbi táblázatban találhatóak [110]:

DOI:10.15774/PPKE.ITK.2023.001

6. táblázat Az egyes aminosavak lehetséges maximális hozzáférhető felszíne

Aminosav	Maximális felszín ($Å^2$)	Aminosav	Maximális felszín ($Å^2$)
Arginin (R)	265	Leucin (L)	191
Aszparagin (N)	187	Lizin (K)	230
Alanin (A)	121	Metionin (M)	203
Aszparaginsav (D)	187	Fenilalanin (F)	228
Cisztein (C)	148	Prolin (P)	154
Glutaminsav (Q)	214	Szerin (S)	043
Glutamin (E)	214	Threonin (T)	163
Glicin (G)	97	Triptofán (W)	264
Hisztidin (H)	216	Tirozin (Y)	255
Izoleucin (I)	195	Valin (V)	165

DOI:10.15774/PPKE.ITK.2023.001

5 Eredmények

5.1 A posztszinaptikus denzitás szerkezeti elemeinek és a betegséget okozó mutációk kapcsolatának vizsgálata

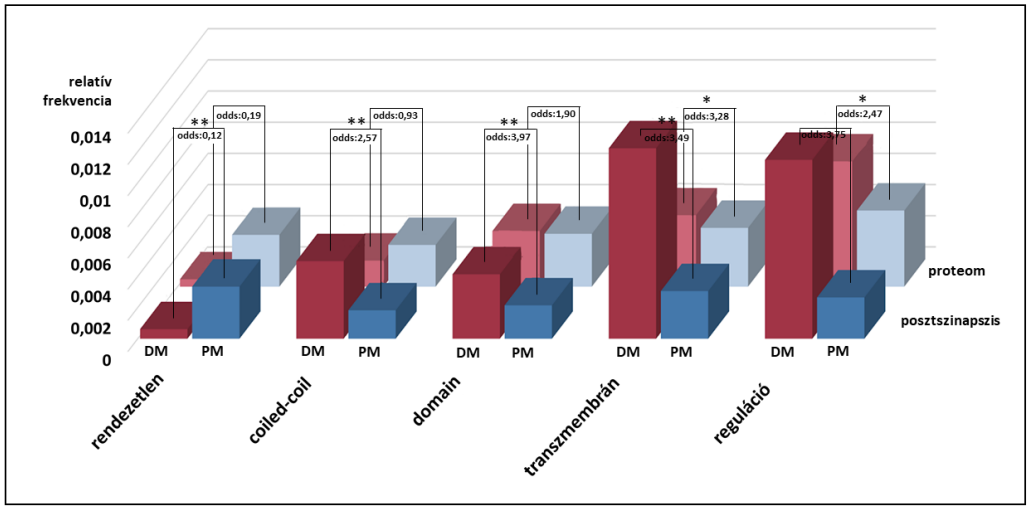
5.1.1 A posztszinaptikus coiled coil fehérjékben jellemzően gyakoribbak a betegséget okozó csíravonal mutációk

A posztszinaptikus (PS) fehérjék szerkezeti kitettségének vizsgálatához meghatároztam a mutációk eloszlását a predikciókkal meghatározott szerkezeti és funkcionális régiókban, mind a posztszinaptikus 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 (7. 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 posztszinapszisban észlelhető trendek ellentétesek: a coiled-coil régiókban a PS esetében magasabb a DM-ek relatív gyakorisága, mint a polimorfizmusoké (7. ábra). Két szempontból is érdekesnek találtuk ezt az eredményt: egyrészt csak ebben az esetben láttunk csak teljesen ellentétes tendenciát a posztszinapszis 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 flexibilis, nagy konformációs szabadsággal rendelkező flexibilis szakaszokon, ahol "nincs szerkezet" ami megváltozhatna, 'elromolhatna'. A jelenség mélyebb megértésének érdekében tovább vizsgáltuk a DM-ek szerepét coiled-coil régiókban. Az ebből származó eredményeket a 5.2. fejezetben ismertetem (7. ábra).

5.1.2 A domének és a rendezetlen szakaszok együttes előfordulása jellemző a DM-kel érintett PS fehérjék esetében

A posztszinapszis struktúrájából következően a fehérjéi rendkívül modulárisak. Ezen megközelítés nyomán a modularitást középpontba állítva elemeztem a DM-ek és PM-ek megjelenését a PS fehérjéiben. Ebben a kontextusban egy modulnak tekintettem bármilyen coiled-coil régiót, rendezetlen szakaszt, domént vagy transzmembrán régiót. Egy fehérjét több modullal rendelkezőnek vettem, amennyiben a felsoroltak közül legalább két különböző szerkezeti elemmel rendelkezett (8. ábra). A alapvető trendek-

DOI:10.15774/PPKE.ITK.2023.001



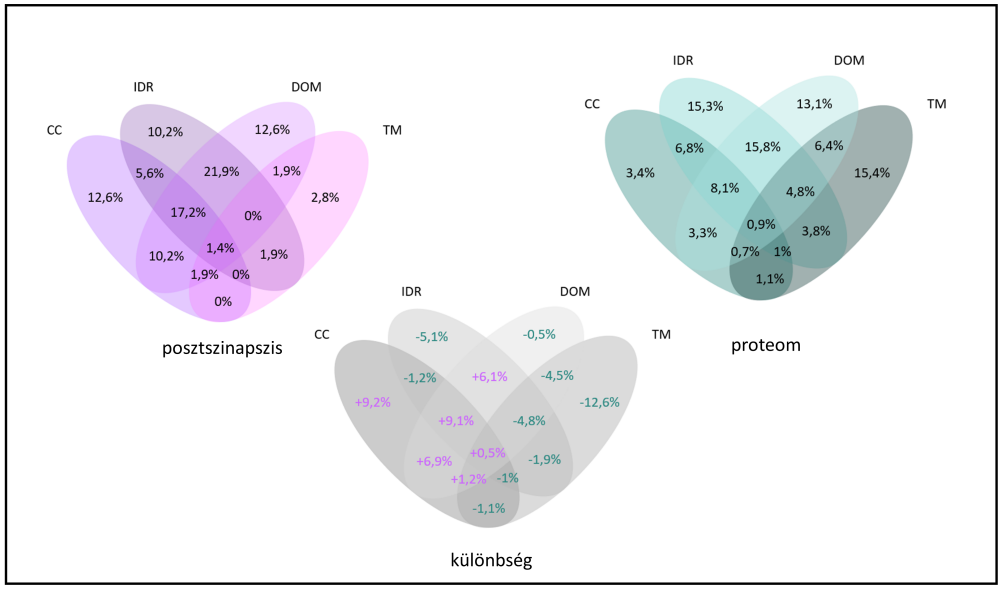
7. ábra A betegséget okozó mutációk (DM) és polimorfizmusok (PM) relatív gyakorisága a PS és a proteom fehérjéiben, az egyes szerkezeti és funkcionális csoportokban (x tengely) A posztszinapsztikus adatszetten számolt frekvenciák esetében a DM vörössel, PM sötétkékkel színezett, míg a proteom esetében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

ben nem látható különbség a PS és a proteom fehérjéi között, mindkét adatszettben az esetek kevesebb, mint felében fordultak elő a mutációk olyan fehérjékben, ami csak egy modullal rendelkeznek. Mind a PS, mind a proteom esetében a jellemző a rendezetlen és domén modulokat tartalmazó fehérjék kitettsége a DM-eknek. Ezek várhatóan két nagyobb fehérje csoportot alkotnak funkcionális szempontból: az egyik esetben a rendezetlen rész linkerként szolgál két domén között, míg máshol maga a rendezetlen rész képes felvenni egy stabil szerkezetet egy partner doménhez kötődve. A két adatszett (PS és proteom) közötti különbség a trendek erősségében érhető tetten. Az egyetlen jelentős különbség közöttük a transzmembrán régiók érintettségében volt, azonban ennek oka valószínűleg a mérési technikákból fakad: a PS fehérjék vizsgálatához (lényegében bármely sejt vizsgálatához) a sejt integritását meg kell bontani, hogy a benne található fehérjék kinyerhetővé és vizsgálhatóvá váljanak. Ennek eredményeként azonban számos, a membránban található fehérje a lipid kettős rétegben ragad és emiatt nem azonosítható.

5.1.3 A posztszinaptikus fehérjék mutációs mintázatának jelentősége

Legjobb tudomásunk szerint a korábban nem elemezték részletesen a posztszinaptikus fehérjék és proteom szerkezeti egységeiben a mutációk dúsulását. Hasonló vizsgálatot azonban végeztek Dobson és munkatársai a teljes humán proteom esetében, ők a globuláris, rendezetlen és transzmembrán adatokon vizsgálták a DM-ek és PM-ek eloszlását [111]. A transzmembrán régiók és a rendezetlen régiók esetében ők is ugyanazokat a trendeket kapták, mint amit mi is tapasztaltunk (a transzmembrán régiókban

DOI:10.15774/PPKE.ITK.2023.001



8. ábra A DM-kel érintett fehérjék modularitásának megoszlása a posztszinaptikus fehérjék és a proteom esetében. A különbséget mutató ábrán a posztszinapszisra jellemző elemek lilával, a proteomra jellemzőbb zölddel (Modulok rövidítései: IDR - rendezetlen, CC - coiled-coil, DOM - domén, TM - transzmembrán)

dúsultak a DM-ek, a rendezetlen szakaszokon dúsultak a PM-ek). Azonban az ő vizsgálataik nem terjedtek ki a coiled-coil régiókra, míg számunkra éppen itt voltak a legérdekesebb eredmények. Első közelítésben a kapott mintázat ellentmond a várhatónak, ennek feloldására az irodalomban próbáltam magyarázatot keresni, azonban a jelenségnek csupán szűk leírását találtam, egyedül Mohanasundaram és munkatársai (2017) [112] foglalkoztak eddig a coiled-coil csíravonalbeli betegséget okozó mutációk általi érintettségével. Ezért a következőkben ezen jelenség bővebb megértésére törekedtünk.

5.2 A coiled-coil szerkezeti elem kitettsége a betegséget okozó mutációk hatásának

5.2.1 A betegséget okozó mutációk ritkábbak coiled-coil régiókban, de gyakoribbak coiled-coilt tartalmazó fehérjékben

A mutációk és coiled-coil szerkezetek kapcsolatának vizsgálatához először egy általános proteom szintű kép felállítása volt a cél a 'CC_SEQ' adatszett-ből kiindulva. A tudományos konszenzussal egybevágóan az látszott, hogy nagy általánosságban a humán proteom fehérjéiben gyakoribbak a polimorfizmusok, mint a betegséget okozó mutációk. Az elemzésből az is észrevehető volt, hogy a DM-ek előfordulása enyhén magasabb volt azokban az esetekben, amikor a fehérje tartalmaz coiled-coil régiót összehasonlítva azokkal az esetekkel, amikor a fehérjében a predikciók alapján nem volt ilyen szakasz (ld. 2. 1. ábra). Ennek a jelenségnek egy lehetséges magyarázata, hogy bár a coiled-coil szerkezeti elemekkel rendelkező fehérjék fontos funkciókat töltenek be a sejtekben - ezért ezek mutációjuknak

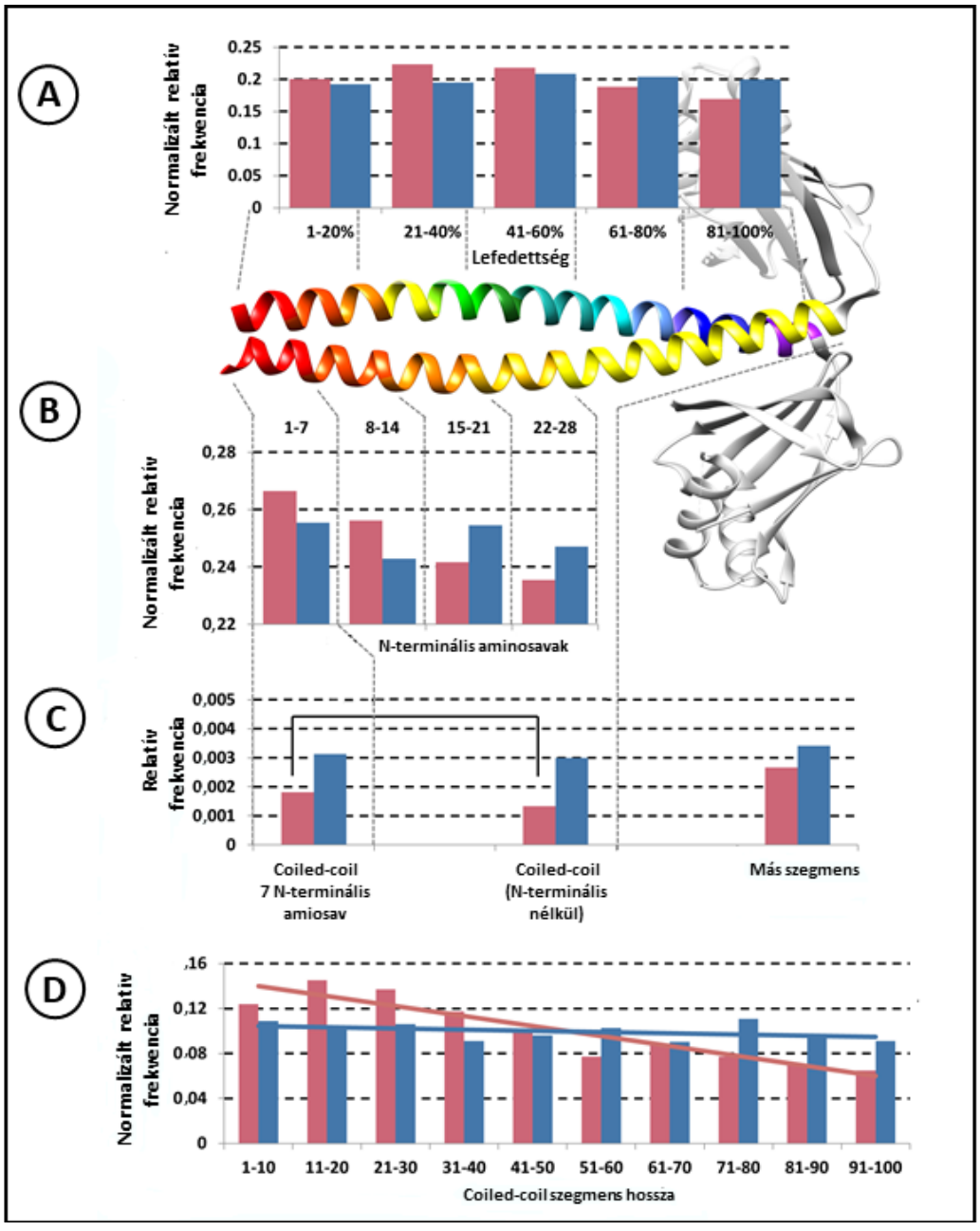
DOI:10.15774/PPKE.ITK.2023.001

komoly fenotípusos következménye lesz -, azonban a coiled-coilon kívül eső régiók sebezhetőbbek. Máshogy fogalmazva, a coiled-coilok mutálódása olyan súlyos következményekkel járhat, amelyek az élettel összeegyeztethetetlenek, ezért nem is lehetséges a továbbadásuk az utódoknak. Hasonló jelenség máshol is megfigyelhető, például a szabályozásban kulcsszerepet betöltő foszforiláció mutációja számok rákos megbetegedés hátterében áll, azonban az örökletes betegségeknél éppen hogy kevésbé jellemzőek letális hatásuk miatt.

5.2.2 A betegséget okozó mutációk feldúsulnak a coiled-coil N-terminális régiójában

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.) (9. ábra, panel A). Ezen megközelítés szerint nem látszódott egyértelműen eltérő dúsulás a különböző régiók között. Ezzel szemben, ha a coiled-coil régiókat a kanonikus heptád ismétlés mentén osztottuk fel (1-7, 8-14, 15-21, 22-28) (9. ábra, panel B). egyértelműen megfigyelhető volt, hogy a DM-ek az N-terminális 7 aminosavat érintik a legerősebben. Az észlelhető különbségek megmutatása mellett ennek a módszernek az előnye, hogy a coiled-coil régiók hosszbeli eltéréseiből származó aránytalanságok is kiküszöbölhetővé váltak. Az N-terminális 7 aminosavban a coiled-coil többi részével összehasonlítva is szignifikáns halmozódást mutattak a betegséget okozó mutációk minden predikció szerint (chi-négyzet teszt, $p < 0.01$). A Bonferroni-teszt alapján korrigálva az eredményt (α korr=0,0025) nem minden egyes predikció eredménye szignifikáns (DeepCoil, Paircoil). Egy megengedőbb, de még mindig szigorú alfa esetén (α=0,05) a korrekció után is szignifikáns az eredmény (α korr=0,0125). Az arányhányados (odds ratio) 1,33 volt a DM-ek és a PM-ek között (9. ábra, panel C). Ez a hatás annyira erős, hogy a rövidebb coiled-coilokban is magasabb a DM-ek relatív gyakorisága, míg a polimorfizmusok hossztól függetlenül teljesen egyenletes eloszlást mutatnak (9. ábra, panel D). Ennek egy lehetséges oka lehet, hogy az N-terminális régiónak kiemelt szerepe van a feltekeredésben. Ismert, hogy a coiled-coil szakaszok feltekeredésében kulcsszerepet játszanak ún. trigger szekvenciák. és kézenfekvő a feltételezés, hogy ez sok esetben az N-terminálison található, bár fontos megjegyezni, hogy a viszonylag kevés, kísérletesen igazolt trigger szekvencia esetében az elhelyezkedés nem mutat ilyen jellegű eloszlást.

DOI:10.15774/PPKE.ITK.2023.001

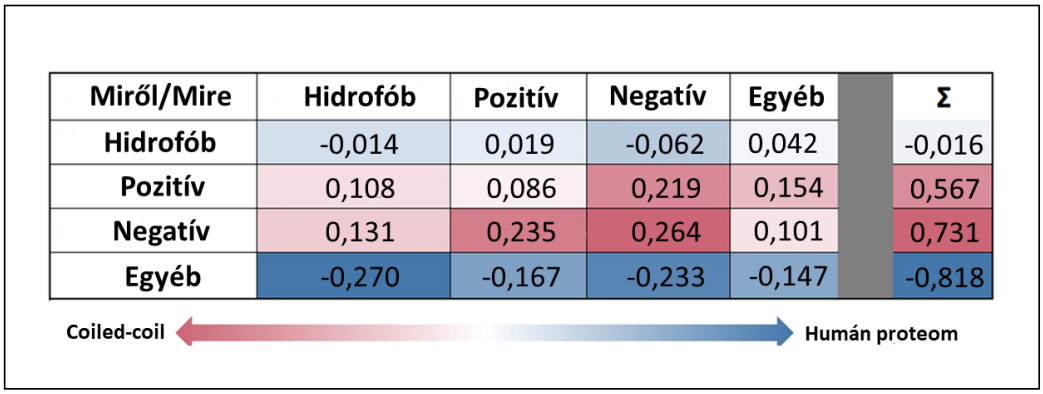


9. ábra A DM-ek és PM-ek eloszlása a coiled-coil szerkezetekben és szerkezeti részeken kívül A: A DM-ek és PM-ek normalizált frekvenciákra a coiled-coil régió hosszában - százalékos felosztás alapján B: A DM-ek és PM-ek eloszlása az regiszterek alapján meghatározott csoportokban C: A DM-ek halmozódása 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)

DOI:10.15774/PPKE.ITK.2023.001

5.2.3 A coiled-coilban történő betegséget okozó mutációk leggyakrabban töltött aminosavakat érintenek

Az aminosavak fizikokémiai tulajdonságait figyelembe véve a betegséget okozó mutációk várhatóan feldúsulnak a stabilitásért felelős kategóriákban. A vizsgálathoz az aminosavakat négy csoportra osztottuk: pozitívan töltött: HKR, negatívan töltött: DE, hidrofób: AILMV és egyéb aminosavak: CFGNPQSTWY. Ennek a rendszerezésnek a coiled-coil szerekezet kialakításában fontos szerepet betöltő aminosavak jellegzetességei adták az alapját (ld. 2.2.2.1. fejezet). A fenti négy kategóriát használva kiszámoltam milyen mértékben változtatják a DM-ek az aminosavakat: (1) a coiled-coilokon belül, (2) a teljes proteomon. A kettő logaritmikus arányát használva láthatjuk, hogy milyen aminosav változások jelentik a preferált DM célpontokat a coiled-coilokon belül. A coiled-coil régiókban a leginkább jellemző cserék a negatívan töltött aminosavak helyettesítése, érdekes módon akár más negatívan töltött aminosavakkal is. Emellett a negatív-pozitív és a pozitív-negatív cserék tekinthetőek még jelentősnek. A humán proteomban preferált cserék az egyéb-hidrofób, illetve az egyéb pozitív töltés voltak. Érdekes módon a glutaminsav(-) és aszparaginsav(-) nem felcserélhető, míg a lizin (+) és arginin (+) cserével sokkal megengedőbb a coiled-coil szerkezet. Ennek oka a glutaminsav hélix szerkezetet elősegítő tulajdonsága [113], amit magányos töltött alfa-hélixek megfigyeltek már [114]. A coiled-coilokban halmozódó DM-ek leginkább az A, E, I, K, L, M, N és Q aminosavakat célozzák, míg a referencia proteomban a C, G, P cserék jellemzőek (ld. Függelék 2. ábra). A hidrofób aminosavak kicserélődésében nem volt nagy különbség a coiled-coil és a referencia között (10. ábra) - ezek az aminosavak mindkét esetben fontos szerepet játszanak a szerkezet összetartásában a hidrofób hatás révén. Az adatokat bootstrap analízis után a 68-95-99.7 szabály alapján vizsgáltam, minden kicserélődés szignifikánsnak bizonyult (ld. 4.4.3. fejezet).



10. ábra Aminosav cserék coiled-coilokban. Bal: aminosav csere preferenciák DM-ekben a proteomban (negatív értékek, kék szín) és coiled-coil régiókban (pozitív értékek, piros szín). Az értékek a DM-ek által megváltoztatott csoportok arányának logaritmusát mutatják a coiled-coil és a proteom esetén. Jobb: Leginkább kitett aminosavcsoportok (negatív, kék: proteom; pozitív, piros: coiled-coil).

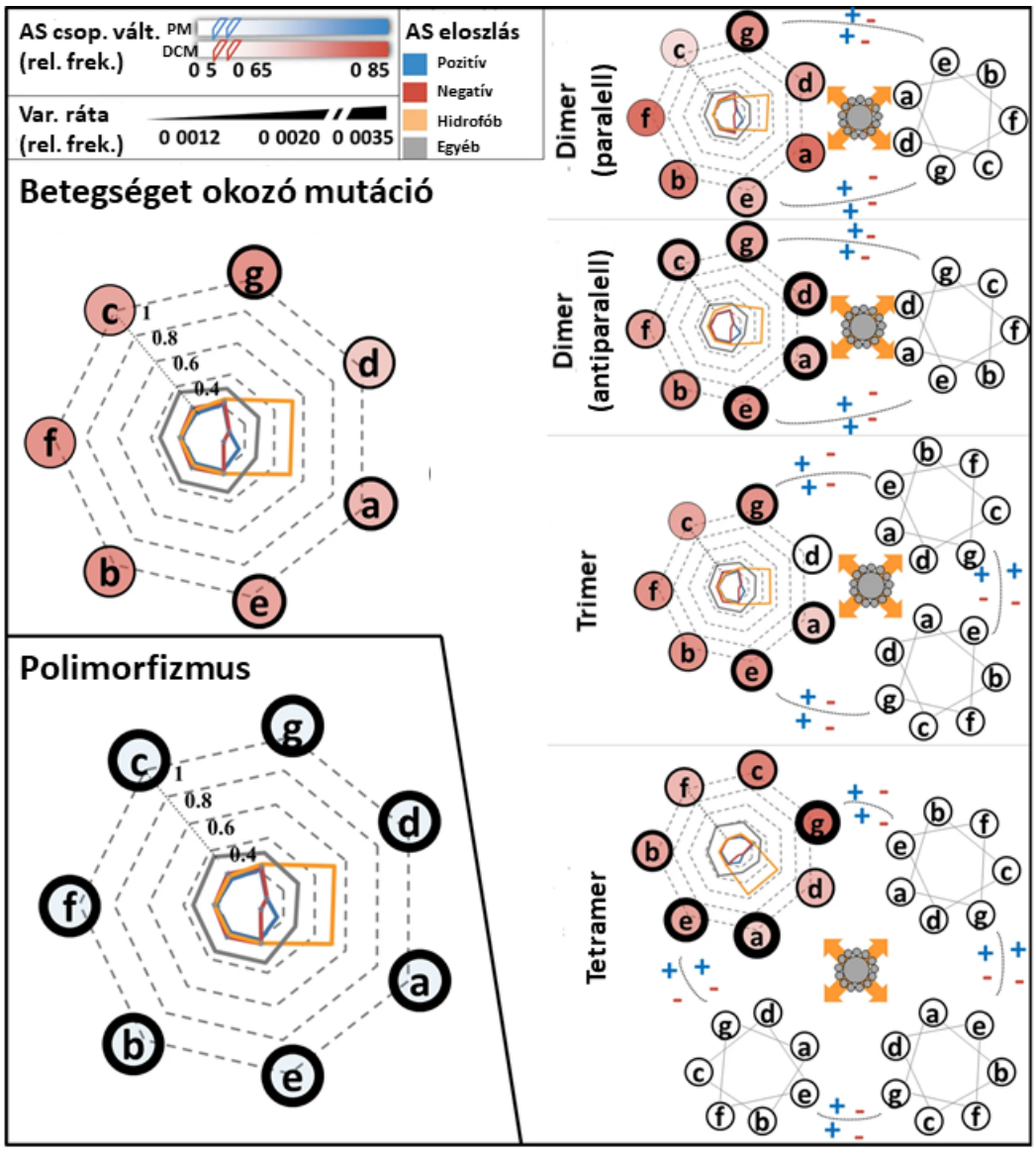
DOI:10.15774/PPKE.ITK.2023.001

5.2.4 Az oligomerizációs állapot befolyásolja a regiszter pozíciók kitettségét a betegséget okozó mutációknak

A következőkben a coiled-coil régiót nem csupán szegmens szinten értelmeztem, hanem a szerkezeti megközelítés felé továbblépve, a heptád ismétlődések figyelembevételével vizsgáltam. A coiled-coil szerkezetekben (ld. 2.2.2.1. fejezet) a struktúra fenntartása miatt bizonyos pozíciókban jellemzően meghatározott fizikokémiai tulajdonságokkal rendelkező aminosavak fordulnak elő. A korábban említett fizikokémiai csoportok alapján az a mintázat jellemző, hogy az ‘a’ és ‘d ‘regiszter pozíciókban hidrofób, az ‘e’ és ‘g’ helyeken a(z ellentétesen) töltött aminosavak előfordulása jellemző, míg a többi pozícióban nagyjából egyenlő eloszlás figyelhető meg a négy csoport aminosavaiból. A mutációkat vizsgálva látható, hogy a DM-ek relatív gyakorisága a coiled-coil szerkezetet összetartó pozíciókban (’a’, ’d’, ’e’, ’g’) a legmagasabb. A ‘g’ pozícióban és az ‘e’ pozícióban jellemző, hogy a DM-ek gyakrabban okoznak fizikokémiai csoport változást, míg az ‘a’ és ‘d’ pozícióban a lévő aminosavak sérülékenyebbek, mivel a csoporton belüli kicserélődést sem tolerálják (11. ábra, bal felül). A polimorfizmusok eloszlása általánosságban egyenletes minden regiszter pozícióban és jellemzően nem járnak fizikokémia csoportok közötti cserével (11. ábra bal alul).

Az egyes oligomerizációs állapotok esetében eltérő halmozódási mintázat látszik. Míg a dimer coiled-coilokban a ‘d’ és az ‘a’ pozíciók azonos mértékben érintettek, addig a trimerekben és a tetramerekben jellemzően magasabb az ‘a’, ‘g’ és ‘e’ pozíciókat érintő DM-ek relatív gyakorisága, de a ‘d’ pozíció kevésbé van kitéve a DM-ek hatásának. A dimerek esetén eltér a parallel és antiparallel coiled-coilokra eső DM-ek relatív gyakorisága: antiparallel esetben az ‘a’, ‘d’, ‘g’ és ‘e’ pozíciók is mutációs célpontok, míg parallel esetben elsősorban a a ‘g’ pozícióra esnek a mutációk. Bár a fentiek pontos magyarázatához kísérletes mérésekre is szükség lenne, összességében megfigyelhető, hogy a DM-ek dúsulása minden esetben a láncok közötti összetartásban szereplő aminosavak esetében volt megfigyelhető (11. ábra, jobb). A kapott eredmények jó összhangban vannak a fehérje tervezési kísérletekben megtapasztalt változásokkal az ‘a’, ‘d’, ‘g’ és ‘e’ pozíciók oligomerizációban betöltött szerepéről [23].

DOI:10.15774/PPKE.ITK.2023.001



11. ábra A mutációk eloszlása a regiszter pozíciókban (általánosan) és a különböző oligomerizációs állapotokban Az ábra bal oldalán az összes coiled-coilon végzett számítás látható, míg jobb oldalon specifikusan, a különböző oligomerizációs állapothoz tartozó coiled-coil-ra jellemzők. A regiszter pozíciók jelölésének közepén található sugár diagramon a négy csoport (poztívan töltött (kék), negatívan töltött (piros), hidrofób (sárga) és egyéb (szürke)) aminosavainak gyakorisága látható az egyes pozíciókban. A regiszeterek körüli vonal vastagsága arányos a pozíciót érintő mutációk frekvenciájával. A körön belüli régió telítettsége azt mutatja meg, hogy mennyire jellemző a fizikokémiai csoportok közötti csere (piros színnel a DM-ek, kékkel a PM-ek vannak jelölve)

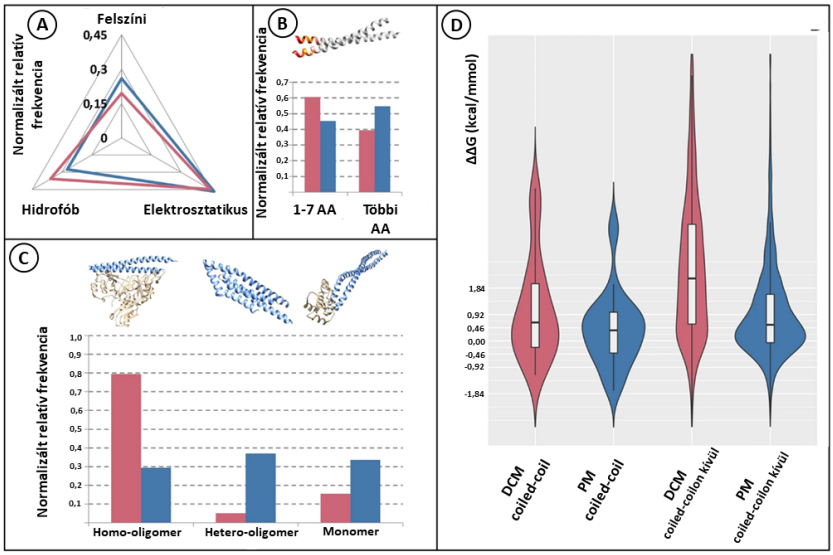
5.2.5 A DM-ek destabilizáló hatással vannak a coiled-coil szerkezetre

A szerkezeti vizsgálathoz a SOCKET eljárás segítségével kerestük a coiled-coil régiókat PDB adatbázisban ('CC_STRUCT' adatszett). A DM-ek és a PM-ek eloszlása a különböző fizikokémiai csoportokban, valamint a DM-ek dúsulása az N-terminális régióban a szerkezeti adatokat vizsgálva is alátámasztást

DOI:10.15774/PPKE.ITK.2023.001

nyertek (12. ábra, panel A és B). A szekvencia alapú predikciók sokszor nehézségekbe ütköznek a coiled-coilok oligomerizációs állapotának meghatározására. A szerkezeti adatok segítségével azonban könnyen és egyértelműen elkülöníthetőek az eltérő oligomerizációs állapotú coiled-coil szerkezetek. A különböző oligomerizációs állapotú coiled-coilok esetében a PM-ek eloszlása a különböző csoportok között egyenletes volt, azonban a DM-ek frekvenciája kiemelkedő a homooligomerekben a heterooligomerekkel és a monomer coiled-coilokkal összehasonlítva (12. ábra, panel C). Ennek magyarázata lehet, hogy ha egy örökölt mutáció megtalálható egy láncon, amely aztán többször előfordul az oligomer szerkezetben, akkor annak hatása megsokszorozódik.

A szerkezeti adatokból a mutációknak a fehérjeszerkezet stabilitására gyakorolt hatása is megbecsülhető. A mutációk hatására bekövetkező energiaváltozást a FoldX program segítségével becsültük. A PM-ek energetikai változása felvázolja azokat a lehetséges eltéréseket, amelyeket a fehérje még nagyobb szerkezeti károsodás nélkül képes elviselni. A DM-ek azonban általánosságban ennél magasabb szabadentalpia változást eredményeznek [111, 115]. A coiled-coilok esetében mind a PM-ek, mind a DM-ek esetében ugyanazt a trendet figyelhetjük meg, mint a proteom esetében, de gyengébb hatásokat látunk: ez alapján úgy tűnik, hogy a coiled-coil régiókba eső mutációk alacsonyabb destabilizáló hatással bírnak, és már így is betegséget okoznak (12. ábra, panel D).



12. ábra A szerkezeti adatokat felhasználva kapott eredmények Az az A-B panelen látható eredmények megerősítették a szekvenciális adatokól levont következtetéseket, A: Az elektrosztatikus és hidrofób kölcsönhatásokban résztvevő aminosavak jobban kitettek, mint a felszíniek B: A DM-ek feldúsulnak az N-terminálison, C: A betegség mutációk leginkább a homo-oligomereket érintik.,D: A coiled-coilba eső mutációk enyhébb destabilizáló hatással bírna (DM: piros, PM: kék)

DOI:10.15774/PPKE.ITK.2023.001

5.2.6 A coiled-coil régiókba eső mutációk sokszor neuronális betegségekkel hozhatóak összefüggésbe

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 DiseaseOntology (DO) ontológia rendszer a humán betegségek leírására készült, ahol az egyes betegségek specifikusan csoportosítva 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. A DO hierarchikus rendszerében az első szint különböző megközelítések szerint osztályozza a betegségeket, amelyek a következők:

- fertőző ágens okozta betegség
- betegségek anatómiai entitás alapján
- sejt proliferációs betegség
- mentális betegség
- metabolikus betegség
- genetikai betegség
- fizikai betegség
- szindróma

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 feldúsulnak izom, szenzoros és bőr betegségek esetében is betegséget okozó mutációk által érintett fehérjék (13. ábra).

DOI:10.15774/PPKE.ITK.2023.001

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

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

5.2.7 Coiled-coil eredmények értelmezése, összevetése korábbi eredményekkel és fehérje biológiai példák

5.2.7.1 Fehérje biológiai példák

A betegségek kialakulásásnak molekuláris szintű megértésétől a legtöbb kórkép esetében még rendkívül messze állunk, léteznek azonban kivételek is, például a 2.4.2. fejezetben említett cisztás fibrózis esetében. Abból kiindulva, hogy a coiled-coil régiók milyen esszenciális folyamatokban résztvevő fehérjékben találhatóak meg, nagy valószínűség szerint számos betegségben érintettek. A következőkben a betegséget okozó mutációk lehetséges szerkezeti és funkcionális következményeit mutatnám be néhány példán keresztül. Szerkezeti megközelítés szerint az aminosavakat érintő mutációk a korábban többször is említett (ld. 2.2.2.1. fejezet) regiszter pozíciók szerint értelmezhetőek. Az 'a' és 'd' pozíciók esetén a sztérikus gátlás és a hidrofób mag megbontása okozhatja a coiled-coil szerkezet sérülését. Az 'e' és 'g' regisztereket érintő mutációk az ellentétes töltésű aminosavak közötti elektrosztatikus kölcsönhatásokat bonthatják meg, illetve bizonyos esetben nem feltétlenül közvetlenül a hélixek közötti kölcsönhatást modulálják, hanem az egyedi lánc hélix szerkezet felvételére való tendenciáját. A coiled-coil szerkezetet összetartó pozíciókon kívül a felületi pozíciók is mutálódhatnak. Ezen három csoportban előforduló aminosav mutációk funkcionális következményét ismertetném a következőkben (az említett sorrendet követve) (14. ábra felső panel). A DM-eket funkcionális szempontból is osztályozhatjuk. Bizonyos esetekben maga a coiled-coil szerkezet sérül, azonban ennek hatása a coiled-coil szakaszon kívül okoz problémát. A desmin-ek olyan citoszkeletális fehérjék, amelyeknek fontos szerepe van a harántcsíkolt izomszövet funkcionális egységének (szarkomer) összehúzódásában. Ezen egységek együttes kontrakciója eredményezi az izomműködés alapját. Strukturális szempontból hosszú coiled-coil régiói kötik össze a desmin fehérje két végén található régiókat, amelyek más fehérjékkel képesek kapcsolódni. Így egy hálózat jön létre, amely összeköti a sejtmagot, a mitokondriumot és izomsejt membránját alkotó szarkolemmát, valamint indirekt kapcsolatban áll az izomsejt alapvető fehérjéivel, az actin-nal és a miosin-nal. A desmin coiled-coil régióban lévő L345P mutáció megzavarja

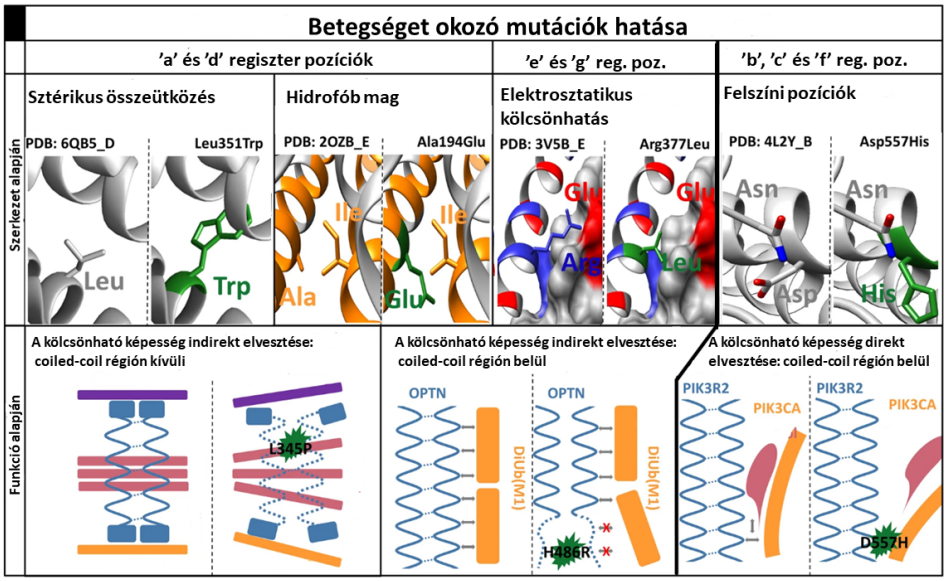
DOI:10.15774/PPKE.ITK.2023.001

az említett hálózat struktúráját: a coiled-coil régió szétesése a rendszer dezintegrációjához vezet, ami végül számos szervre (többek között szívre) kiterjedő komoly betegséget okozhat, ezen kórképek összefoglaló neve a dezminopátia (14. ábra, alsó panel bal)[116]. Egy másik lehetséges funkcionális következmény, amikor a coiled-coil stabilitásáért felelős pozíciójában történő mutáció hatására a coiled-coil kívülről is hozzáférhető részén hiúsul meg egy kölcsönhatás. Az optineurin egy állványfehérje, amely autofágia receptorként működik. Az autofágiára szánt "szállítmányt" képes összekapocsolni az ubikvitinációt végző fehérjékkel. Az optineurin ubikvitin kötő régiója egy hosszú coiled-coil régióban található. A H486R mutáció jelentősen befolyásolja az optineurin és ubikvitináz kapcsolódását, amely egyes autofágia folyamatok megváltozását okozhatja. Az optineurin előbb említett mutációját a glaukóma (visszafordíthatatlan vakság) egyik altípusának kiváltójaként is leírták (14. ábra, alsó panel középen)[117]. Más esetben a coiled-coil víz (és más molekulák) számára is hozzáférhető részén történik egy mutáció és ez direkt módon megváltoztatja a coiled-coil szakasz egy másik fehérjével való kölcsönhatását. A PIK3R2 a phoszphatidil-inozitol 3-kinase enzim egyik alegysége. A PIK3R2 a PIK3CA-val heterodimer szabályozó egységet képez, a kapcsolódásukban kulcsszerepet játszik a PIK3CA N345 aminosava, és a PIK3R2 D557 aminosava közötti hidrogénhíd kötés. A két fehérje által létrejövő kapcsolat egy zsebet hoz létre, amely fizikai teret biztosít a glicerin megkötéséhez. A PIK3R2 557-es aminosavnak esszenciális szerepe van a zseb kialakításában, amit az Asp ->His csere képes megzavarni, mivel az Asp oldalláncának a glicerinnel való közvetlen kapcsolata így elvész, valamint a molekulát pozicionáló negatív töltés hiánya (amely a pozitív töltésű és nagyobb hisztidin hatására megszűnik) negatívan befolyásolja a kötődést. A PIK3 számos esszenciális jelátviteli útvonal fontos sze-replője. A konkrét mutációt egy komplex agyi rendellenességet, értelmi fogyatékosságot és összetett részleges rohamokat mutató betegnél mutatták ki (14. ábra, alsó panel jobb) [118].

5.2.7.2 A coiled-coil régiók mutációs mintázatának értelmezése és összevetése korábbi adatokkal

A predikciókon alapuló bioinformatikai elemzéseknél az egyik legfontosabb kérdés az így létrehozott adatok megbízhatósága. Ez növelhető több predikció konszenzusának használatával, vagy azok eredményeinek kritikus összevetésével. Azért, hogy elkerüljük az egyedi módszerek sporadikus hibáit, mi is több módszert használtunk a coiled-coil régiók meghatározására és a regiszterek definiálására. A coiled-coil predikáló módszerek megbízhatóságát külön nem elemeztük, ugyanakkor Simm és munkatársai friss eredményei alapján , amelyben az akkor elérhető szerkezeti adatok alapján értékelték a coiled-coil predikciók megbízhatóságát és nem találtak jelentős különbséget: a különböző módszerek pontosságát tekintve, az Ncoilst (∼70%) leszámítva mindegyik módszer 80%-os feletti pontossággal határozta meg a coiled-coil régiókat a fehérje szekvenciákból. Ez nagy mértékben összhangban van az egy évtizeddel korábban Szappanos és munkatársai által publikált eredményekkel is [119]. Fontos kiemelni, hogy a DeepCoil predikciós algoritmust nem vették figyelembe, pedig ez a jelenleg elérhető legújabb

DOI:10.15774/PPKE.ITK.2023.001



14. ábra A coiled-coil mutációk szerkezeti és funkcionális következményei a regiszterek DM érintettsége alapján A mutációk szerkezeti megnyilvánulása (felső panel): 'a' és 'd' pozíciók esetében sztérikus gátlás vagy a hidrofób mag megbontásában játszik szerepet a DM (bal); az 'e' és 'g' pozíciók esetében az elektrosztatikus kölcsönhatás megzavarása (középen); a 'b', 'c' és 'f' pozíciók esetében a felszín érintettsége A mutációk funkcionális következmények (alsó panel): interakciók megzavarása: a coiled-coil régión kívüli részen (bal) a coiled-coil régión belül, indirekt módon (közép) módon és a coiled-coilon belül, direkt módon (jobb)

módszer, ráadásul ez már a kor elvárásainak megfelelően mély gépi tanulást alkalmaz. A coiled-coil régiókon végzett számolásokat külön-külön elvégezve a kapott eredmények egybehangzóak voltak (az ábrákon minden esetben a módszerek átlaga látható). Az adatok megbízhatóságát az is alátámasztja, hogy a főbb állításokat a szerkezeti adatokon végzett számítások is megerősítették, azok eredménye konzisztens a szekvencia elemzéssel kapottakkal. Ahogyan korábban is hivatkoztam rá, a betegséget okozó csíravonal mutációk hatásait egyetlen tanulmány vizsgálta eddig nagyskálásan coiled-coil szakaszokon. Mohanasundaram és munkatársai más megközelítésből analizálták a jelenséget. Esetükben a fő hangsúly az irregularitásokon és a pleotrópián volt, ezzel szemben az elemzésük sokkal általánosabb volt. Emellett a megbízhatóság növelése érdekében mi négy módszerrel határozzuk meg a coiled-coil régiókat, szemben az általuk egyedül használt Marcoil predikcióval. A regiszterek kitettségét mindkét tanulmány elemezte és hasonló eredményeket kaptunk.

DOI:10.15774/PPKE.ITK.2023.001

5.3 A PostSynapticInteractionDataBase (PSINDB) felállítása

5.3.1 A PSINDB adattartalmának és struktúrájának meghatározása

A coiled-coil, mint interakciós motívum kitüntetett szerepe a posztszinapszis szerveződésében még inkább ráirányította figyelmünket a fehérje-fehérje kölcsönhatások szinaptikus jelátvitelben játszott szerepére. Áttekintve az elérhető adatforrásokat, arra a következtetésre jutottunk, hogy bár léteznek szinaptikus fehérjékre specializált adatbázisok, és számos általános fehérje-fehérje interakciós (protein-protein interaction, PPI) adatkészlet elérhető, ezek a gyakorlatban nem használhatóak igazán hatékonyan a posztszinapszis fehérje hálózatának célzott vizsgálatában. Ennek okai között a PS fehérjék általános adatbázisokban lévő hiányos annotációja, valamint az egymással kölcsönhatásba lépő fehérjék PS-beli lokalizációjáról való ismeretek bizonytalansága is kiemelhető. Igen fontos aspektusként jelent meg továbbá a PS-re jellemző nagyszámú multivalens interakció reprezentálhatósága miatt a kötőrégiók lehetőség szerinti minél részletesebb ismerete. Mindezek miatt elhatároztuk egy új, kifejezetten a posztszinaptikus fehérje-fehérje interakciókra specializált adatbázis létrehozását. Az alapvető célunk az adatbázis felállításánál egy olyan posztszinapszis specifikus bináris fehérje-fehérje kölcsönhatásokat 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 (ld.2.3.4. fejezet). A PSINDB-ben két információra kereshetünk, és ez a két információ adja az alapját az adatok megjelenítésének is. Az első a fehérjék adatlapja, ahol a 14. ábrán megjelölt összes információ megjelenik. A másik a bináris interakciók oldala, ahol két fehérje kölcsönhatása látható. A fehérje információs oldal a következő bekezdés szerint meghatározott, jelenleg összesen 2 160 posztszinaptikusnak számító fehérjére érhető el. Részletesebben az oldalak megjelenéséről a 5.3.3. fejezetben lesz szó. A PSINDB felállításához első lépésben a posztszinaptikus adatszett meghatározására volt szükség. A korábbi szűk értelmezés és szigorú kritériumokkal szemben a mostani esetben az inkluzivitást tartottuk szem előtt. A PSINDB mellett a területen meghatározó, új adatbázis a SynGO, amelyet 2019-ben publikálták, és amelynek adataira mindenképpen fontos elemként tekintettünk. A PS fehérjéit definiáló adatszettet ezért a PSINDB esetében kiegészítettük a korábban használt SynaptomeDB és G2C mellett a SynGO adatbázissal, valamint figyelembe vettük a GO posztszinaptikus lokalizáció annotációval ellátott fehérjéit is. Az ebből a négy adatbázisból származó adatok mellett a Human Protein Atlas alapján a neuronális expressziót is figyelembe vettük (15. ábra, panel A). Az adatbázisban az egyedi fehérjékre vonatkozó adatok (15. ábra, panel B-F) csak a posztszinaptikus fehérjékre találhatóak meg, azonban az interakciós partnerek listájában nem PS fehérjék is megjelennek. Ennek oka azon korábbi felvetés, hogy a szinapszis komplex és eltérő összetétele miatt valószínűleg nem minden szinaptikus fehérjét sikerült még azonosítani. Ugyanakkor ezáltal az adatbázis egyértelműen elkülöníthetővé teszi az ismert posztszinaptikus lokalizációval rendelkező

DOI:10.15774/PPKE.ITK.2023.001

fehérjék között kialakuló, várhatóan a PS felépítésében is releváns komplexeket az esetlegesen egyéb szövetekben előfordulóktól. 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), fázisszeparáció (PhasePro), lineáris motívumok (ELM), foszforiláció (UniProt annotáció), doménnek (PFAM). Az interakciós adatok esetében a manuális kiértékelés mellett négy adatbázis adatainak integrálása történt meg: IntAct, BioGRID, STRING, valamint a PDB-ben található több alegységes fehérjék kölcsönhatásai. Az adatbázisba humán, egér és patkány adatok kerültek be, minden esetben visszatérképezve a humán ortológra (OMA). A négy adatbázis eltérő mélységű adatot tartalmazott, azonban minden esetben a MIMIx+ rendszer előírásai szerint jelenítjük meg az adatokat (a nem ismert információkat külön jelöljük). Emellett az interakciók fel is vannak címkézve, evidencia alapján, ami lehet:

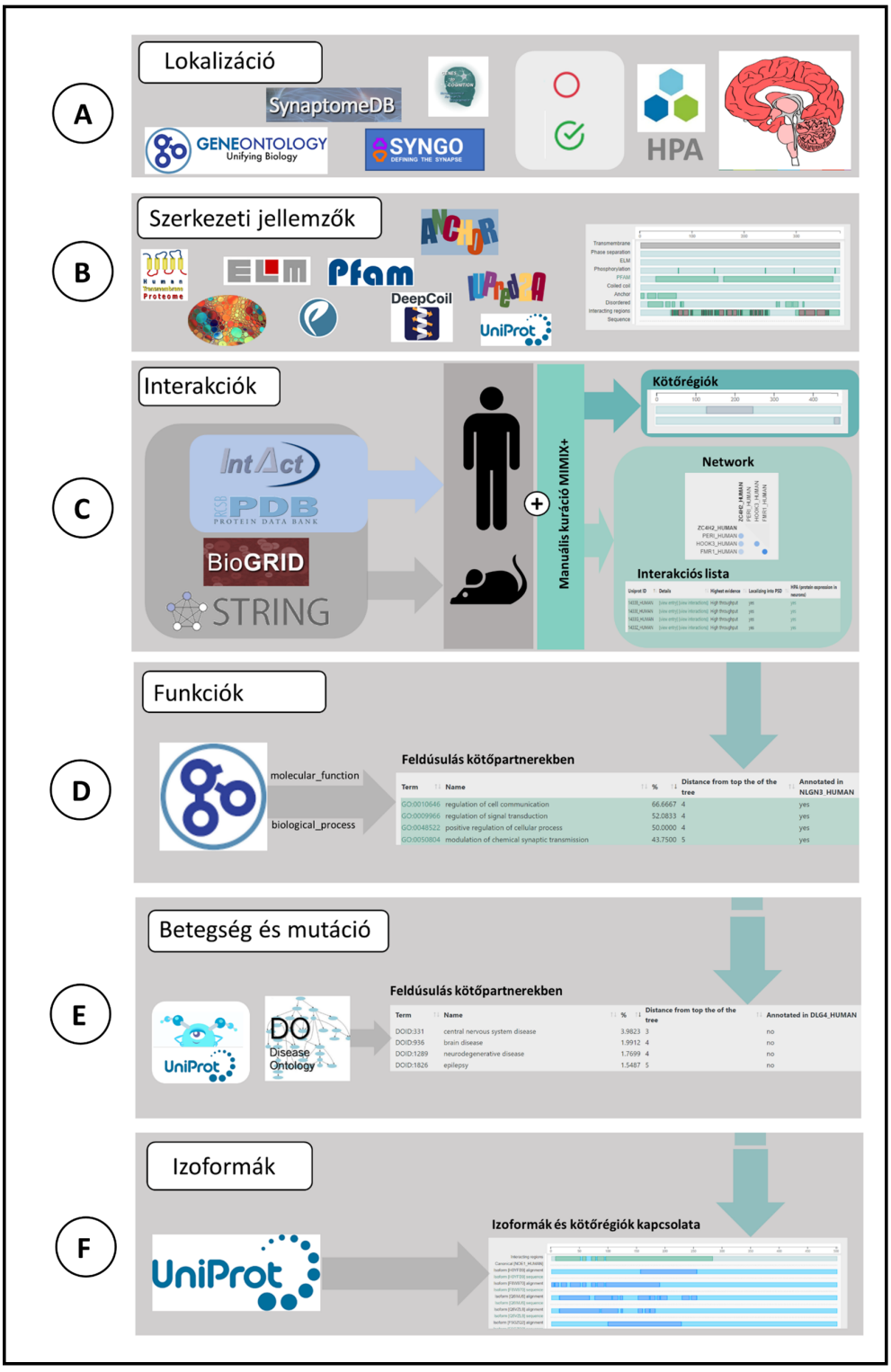
- alacsony áteresztőképességű kísérlet (low throughput)
- nagy áteresztőképességű kísérlet (high throughput)
- számításos predikció (computational)

Az adatbázis létrehozásakor különös figyelmet fordítottunk arra, hogy a kötőrégiókat minél részletesebben megjelenítsük. Kötőrégiók három forrásból származnak: az IntAct adatbázisból, saját annotációból irodalmi adatok alapján, valamint a PDB adatbázisbeli szerkezeti információkból. Utóbbi esetben minden posztszinaptikus fehérjét tartalmazó bejegyzésben azonosítottuk az egymással kölcsönhatásban lévő aminosavakat a Voronota programmal. Ilyenkor a PDB-ben mindig a PISA által meghatározott legvalószínűbb oligomerizációs szintet vettük figyelembe. Az ortológ fehérjékről átvett adatok esetében az interakció tényét rögzítjük, azonban a régiókat nem vesszük át. Ennek oka, hogy a lineáris motívumokat érintő interakciók esetében ezen motívumok sajátosságai miatt a konzerváltságon alapuló megfeleltetés még az egyes ortológ fehérjék között sem tekinthető megbízhatónak, és nem is mindig triviálisan kivitelezhető. Mivel egy kölcsönhatásról több kísérletből is rendelkezésre állhat információ a kötőrégióról, ezeket mindig összesítettük. A kötőrégiók meghatározása több szinten történhet a HUPO-PSI szerint: binding associated region, sufficient binding region, necessary for binding, direct binding. Azonban ez utóbbira nagyon kevés példa volt (<10), nem került bele az adatbázisba, viszont a PDB-ből származó adatokat figyelembe véve hozzáadtuk az ‘atomic’ szintet, ami a ténylegesen atomi szintű kölcsönhatást jelöli 2 polipeptidlánc között. A szekvenciális megjelenítés mellett a UniProt webes felületéhez hasonlóan “hálózatosan” is ábrázoljuk az interakciókat, így jól áttekinthetővé válik, hogy egy nagyobb hálózat mely elemei állnak egymással kapcsolatban. Ez az információ listaszerűen is megjelenik, így a különböző partnerek oldala egyszerűen elérhető (15. ábra, panel C). Mivel számos posztszinaptikus fehérje funkciója még nem ismert részleteiben, úgy gondoltuk, hogy érdekes lehet

DOI:10.15774/PPKE.ITK.2023.001

a STRING-ben megtalálható módon funkcionális “ujjlenyomatot” adni a fehérjékhez kötőpartnereik alapján. Ehhez a GO molekuláris funkció és biológiai folyamat terminusok százalékos megjelenését mutatjuk a fehérje interakciós partnerei között, és megjelöljük azokat, amelyek az adott fehérjéhez is hozzá vannak rendelve (15. ábra, panel D). Ezáltal tulajdonképpen egy komplex deszkriptort rendelünk a fehérjéhez, amely pl. hasonlósági keresésekben is használható lehet. Hasonló elv mentén a DiseaseOntology alapján a partnerekben dúsuló betegség csoport terminusok is megtalálhatóak. Mivel az ontológiák hierarchikusan rendezettek, az egyes kifejezések sorbarendezhetőek specificitás alapján (pl. egy magasan lévő kifejezés lehet nagyon általános, de több fehérjére fog illeni). A PSINDB-ben ezen kívül megjelenítjük a különböző ismert csíravonalbeli mutációkat a Humsavar alapján, és a részletes annotáció jóvoltából az is látható, ha egy mutáció ismert kötőrégióba esik (15. ábra, panel E). Egy fehérjének több izoformája létezik, ez a szabályozás egy további szintjét jelenti sok fehérje esetében. Az izoformákból gyakran éppen egyes kötőrégiók vágódnak ki, így érdemes ezeket az adatokat is összevetni. A UniProtban található izoformákat is átvesszük, minden szekvencián jelöljük a kivágódó szakaszokat, referenciaként pedig az összesített kötőrégiókat is mutatjuk (15. ábra, panel F).

DOI:10.15774/PPKE.ITK.2023.001

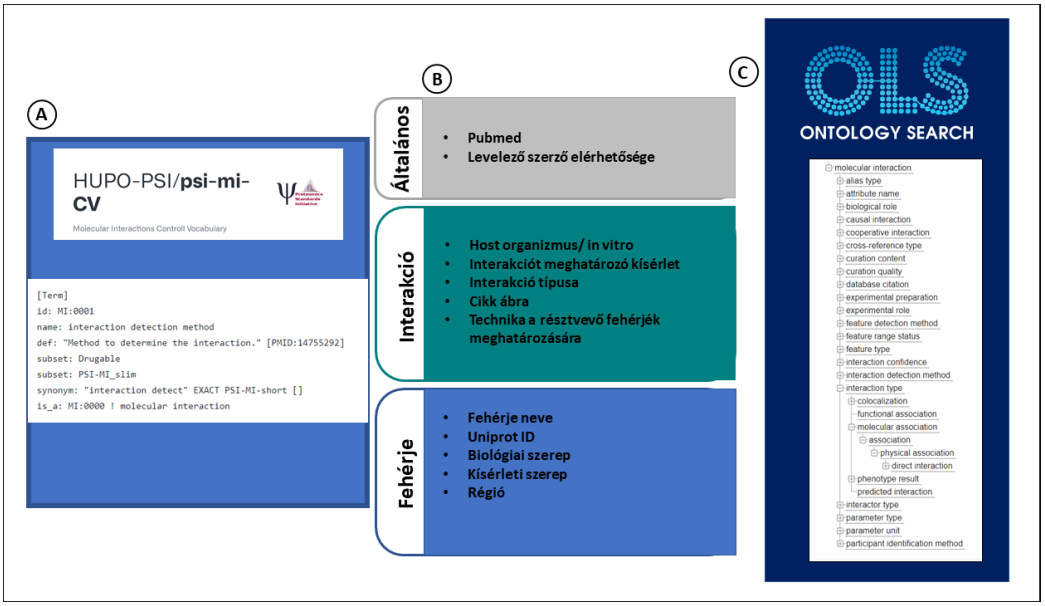


15. ábra A PSINDB adatrétegei A: A posztszinaptikus lokalizáció (forrás: SynaptomeDB, G2C, SynGO, GO) és expresszió (forrás HumanProteinAtlas), B: A fehérjék szerkezeti és funkcionális jellemzői (főbb források: IUPred2A, PFAM, UniProt, részletekért lásd a szöveget), C: Interakciók, D: Funkciók, E: Betegség és mutáció, F: A fehérje izoformák és kötőrégiók kapcsolata

DOI:10.15774/PPKE.ITK.2023.001

5.3.2 Az interakciók feldolgozási folyamatának és reprezentációjának meghatározása

A PSINDB létrehozásakor egy eddig nem említett, de fontos törekvés volt a terület legjobb gyakorlatainak megismerése, átvétele és alkalmazása. Ehhez a HUPO (Human Proteome Organisation) Proteomics Standards Initiative Fehérje Interakciós csoportjának javaslatait követtük. Emellett azt is fontosnak tartottuk , hogy ne növeljük a redundanciát a publikációkból, ezért a manuális feldolgozás előtt a cikkeket az IMEX central rendszerében ellenőriztük, és az ott nem szereplő publikációk kerültek a következő, manuális kurációs körbe. Ahogy korábban a 2.3.4. fejezetben már említésre került, az interakciók leírására a 'gold standard'-nak számító eljárás az IntAct szintű mély reprezentáció. Ez bizonyos kísérlettípusok esetében, mint a röngten-krisztallográfia vagy az NMR, viszonylag egyértelmű és belátható folyamat (az adatbázisunk automatikusan tartalmazza is ezeket), azonban a legtöbb kísérletes technika esetében rendkívül erőforrásigényes, és sokszor nehézségekbe ütközik a precízleírás. 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 kötő régiók is rögzítődnek. A HUPO-PSI ajánlásokat más területen is követtük. A kísérletek és azzal összefüggő információk leírásához a PSI-MI kontrollált szókészletet (Controlled Vocabulary, CV) és a kereséshez az EBI ontológiakereső oldalát, az Ontology Lookup Service-t (OLS) használtuk. Végül, de nem utolsósorban, a PSI által ajánlott PSI-MI (Proteomics Standards Initiative, Molecular Interactions) formátumban tettük közzé az adatainkat.



16. ábra A MIMIx plusz adatok (középen), és az interakciók leírása során felhasznált terminus leírások gyűjteménye (HUPO-PSI-MI-CV)(bal) és keresési lehetőségük (Ontology Lookup Service)(jobb)

DOI:10.15774/PPKE.ITK.2023.001

Az annotálás folyamata az alábbiak szerint zajlott: a MIMIx (+) rendszerben három információ halmazba tartoznak az adatok: általános, ahol a PubMed azonosító van megadva; interakciós adatok, amelyek jellemzően a kísérlet körülményei; a harmadik csoport magára a fehérjére/fehérjékre vonatkozik (16. ábra). Általánosságban egy annotációs folyamat így írható le:

1. PubMedID meghatározása
2. A cikk átolvasása, elsősorban az ábrák segítségével, illetve az absztrakt elolvasásával
3. Kísérletek azonosítása olyan szempont szerint, hogy egyértelműen meghatározhatóak legyenek a bennül szereplő fehérjék - ábrák, szöveg és módszerek figyelembevétele A kísérletben kimutatott interakció helyének meghatározása: *in vivo*/*in vitro*, sejtvonalak
4. Fehérjék meghatározása és azonosítók kigyűjtése
5. Kísérlet és abból következően az interakció típusának meghatározása
6. Fehérjék kísérleti és biológiai funkciójának elemzése
7. Kötőrégió meghatározása a UniProtID alapján az ott található szekvenciával összevetve

Ahogy a 2.3.4. fejeztben már hivatkoztam rá, az annotáció egyik legnagyobb kihívást jelentő része a kísérletben résztvevő fehérjék meghatározása, úgy hogy egyértelműen hozzárendelhetőek legyenek egy adatbázisazonosítóhoz. Sok esetben előfordul, hogy egy laboratórium évek óta végez kísérleteket egy adott fehérjén, ezért a kísérleti konstrukt, amit használnak, egy korábbi cikkben került leírásra. Ilyenkor a meghivatkozott cikket is a fenti pontok alapján elemezni kell, azért hogy minden bizonyossággal állítható legyen, mely fehérje szerepelt a vizsgálatban. Nagy múltú laboratóriumok (C. Sala csoportja, CNR Neuroscience Institute, Milano) esetében ez a láncolat kettőnél több lépésből is állhat. Emellett a fehérje izoformák is megnehezítik az azonosítást. A megfelelő izoforma azonosításhoz a fehérje szekvenciát is látni kell, amely azonban nem minden cikkben szerepel. Az interakcióba lépő fehérjék meghatározási nehézsége miatt, számos értékes adatot kell eldobni, amely így nem kerülhet be adatbázisokba. A interakciós halmazba tartozó adatok esetében is nehézségekbe ütközhetünk, azonban ezek teljesen más jellegűek, mint a fehérjékkel összefüggő adatok. Például a szakirodalomban, és sokszor publikációkon belül is következetlenül használják a kísérletes technikák megnevezését. Emellett fontos azt is megjegyezni, hogy kísérletek esetében a tudományos bizonyíték a cikkekben közölt ábrák, amelyek értelmezése része a kurációs folyamatnak. Gyakran utólag bizonyosodik be, hogy a cikkben közölt ábra és a szöveg nincs összhangban, vagy rosszabb esetben manipulált, ahogy az nemrég az Alzheimer-kórral kapcsolatos kutatásokat is megrengette. Bizonyos esetekben véletlen hibák vagy akár szándékos visszaélések érhetőek tetten a közölt ábrákon, azonban ezeket sokszor még szakavatott szemeknek is nehéz észrevenniük. Az adatbázisok egyik fontos tulajdonsága, hogy mennyire megbízható

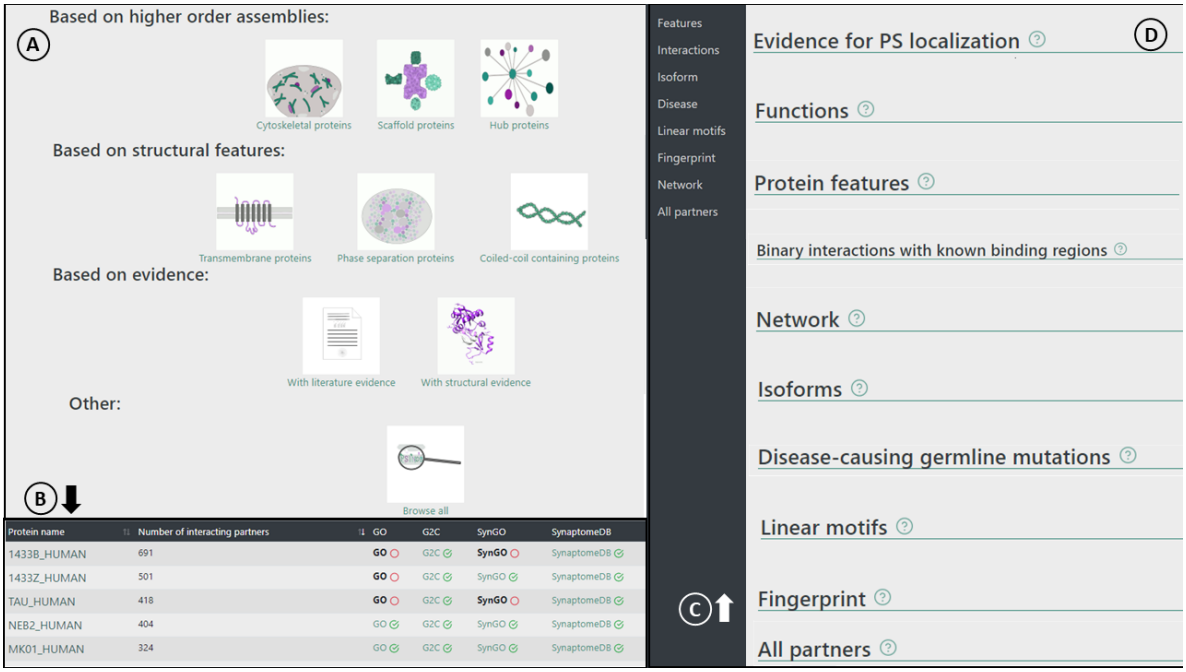
DOI:10.15774/PPKE.ITK.2023.001

adatokat tartalmaznak. A PSINDB megközelítése (hasonlóan az IntAct rendszeréhez) az az elv, miszerint az eredményeket csak közzétesszük, szakmai kritikát az eredményekről nem fogalmazunk meg, annak megítélését a végfelhasználóra bízzuk, hogy megbízik-e a felsorolt bizonyítékokban. A manuális kuráció végeredményeként több, mint kétezer új kísérletes bizonyíték került az adatbázisunkba, ami eddig nem volt rendezetten hozzáférhető, és szabadon is letölthető.

5.3.3 A PSINDB technikai megvalósítása és felhasználói felülete

Az adatbázis magját egy MySQL adatbázis adja, amelyet Dobson László hozott létre, a weboldal felhasználói felületének megvalósítása Dudola Dániel munkája, amelyhez Django keretrendszert használt. Az adatbázis adatokkal való feltöltése teljes egészében a saját munkám, valamint természetesen részt vettem az SQL szerkezetének és a felhasználói felület felhasználói élményének (UX design) a megtervezésében is. Az adatbázis sémája a Függelék 3. ábráján látható. Az adatbázis fizikailag a PPKE ITK szerverein található, elérhetősége:
`https://psindb.itk.ppke.hu/`.
Az adatbázis weboldaláról letölthetőek a kísérletesen meghatározott kölcsönhatások, a posztszinaptikus fehérjék listája, a posztszinaptikus kölcsönhatási hálózat, valamint a kölcsönhatásokban részt vevő régiókat tartalmazó fájl. A weboldalon a rákereshetünk fehérjékre UniProt azonosító, névalapján vagy génnév segítségével, illetve külön a kölcsönhatásokra is. A jobb felhasználhatóság miatt egyes általunk előre definiált fehérjecsoportok is egyszerűen áttekinthetőek (17. ábra, panel A). A fehérjecsoportokat úgy határoztuk meg, hogy a PS szempontjából relevánsak legyenek: citoszkeletális fehérjék, fázisszeparációban részt vevő fehérjék, transzmembrán fehérjék stb. Az egyes csoportokon belül a fehérjék megjelenítése látható az ábra B panelén. A fehérjéket nevük alapján abc sorrnedbe, illetve az interakciós partnereik száma alapján is rendezhetjük. A kölcsönhatások száma mellett a posztszinaptikus lokalizáció forrása is megjelenik (17. ábra, panel B). Az egyedi fehérje lapján oldal panel segítségével is navigálhatunk (17. ábra, panel C), illtve a fehérje oldalán legörgetve is elhatároljuk a különböző információkat (17. ábra, panel D).

DOI:10.15774/PPKE.ITK.2023.001



17. ábra A tájékozódást segítő elemek a PSINDB-ben A: Fehérje csoportok szerinti keresés B: Csoporton belüli keresés - fehérje listák (rendezés A-Z, interakciós partner szám alapján), C: Fehérje oldalon belüli keresés D: Az oldalon belüli szekció jelölések

5.3.4 Esettanulmányok a PSINDB adatbázis használatára

A következő példák segítségével szeretném bemutatni, hogyan szolgál a PSINDB adatbázis, mint egy központi adatforrás a posztszinapszis vizsgálatához. A bemutatott példák a PSINDB adatintegrációján alapulnak, így más adatbázis vagy szerver segítségével nem valósíthatóak meg. Ezek a felhasználási módok lehetővé teszik a PS fehérjék egyedi és nagyskálás vizsgálatát is. A PSINDB felhasználható lokális alhálózatok részletes vizsgálatára: a SHSA6 és SHSA7 fehérjék (18. ábra) 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 [120]. A UniProt funkcionális leírása szerint a hippokampusz C1, C3 régióinak szinapszisaiban töltenek be szerepet a transzmisszió fenntartásában, 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. Mind az SHSA6, mind a SHSA7 fehérjében nagy valószínűség szerint olyan motívumok találhatóak, amelyekek 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

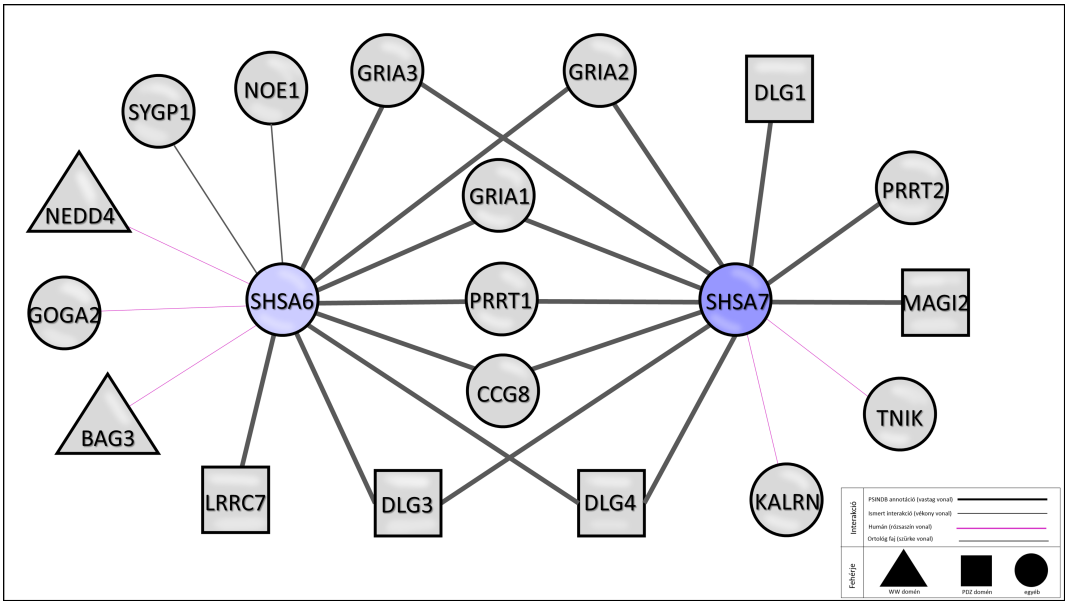
DOI:10.15774/PPKE.ITK.2023.001

[121]. Az újonnan feltárt partnereik között öt fehérjében található meg PDZ domén. Emellett nagy valószínűséggel tartalmaznak WW domén kötő motívumot is (ELM predikció alapján), 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 [122], illetve idén közöltek egy tanulmányt, amelyben kimutatták a fehérjét kódoló gén érintettségét kognitív zavarban [123]. Az PSINDB adataira támaszkódó elemzés eredményeként egy jóval komplexebb képet kaptunk ezen fehérjék neuronális szerepéről. Ez a példa is jól mutatja, hogy az adatok integrálása egy adatbázisban segítheti a rendszerek bővebb megértését. Emellett rávilágíthat a rendszer azon pontjaira, ahol a jelenlegi adatok tovább bővíthetőek a kísérletes adatok gyűjtésével és rendszerezésével.

A PSINDB-ben számos PS fehérje részletes annotációja található meg: a Shank fehérjecsalád tagjai az egyik legfontosabb állványfehérjék a posztszinaptikus denzitásban [124]. Ezek a multidomén fehérjék összekötik a membránba ágyazott receptorokat a citoszkeleton rétegeivel. A Shank fehérjék polimerizációra is képesek a SAM doménjükön keresztül [125], valamint hálózatot alkotnak a Homer fehérjékkel [126]. A Shank fehérjékben előforduló betegséget okozó mutációk számos neurodegeneratív betegséggel összefüggésbe hozhatók [127]. A Shank3 fehérje 280 partnerrel rendelkezik a PSINDB adatbázisban, amiből 77 esetben a kötőrégió is meg lett határozva. Például az Abi1 a Shank3 374-739 valamint a 791-1221 szegmensével is kölcsön tud hatni, míg a gapdh a 374-739 és a 1219-1730 szakaszokkal léphet kapcsolatba: feltételezhetően a Shank3 konformációs változásai a fehérje más és más részeit teszik hozzáférhetővé, hasonlóan ahogy az autizmussal összefüggésbe hozható mutációknál leírták [128]. A PSINDB-ben található részletes kötőrégió adatok megkönnyítik a PS komplexeinek modellezését, ahogy ezt Miski és munkatársai is bemutatták [129].

A PSINDB-ben alapvetően a PS-ben található fehérjék kölcsönhatásaira koncentráltunk, azonban az adatbázis tartalmaz minden olyan egyéb interakciót is, ahol csak az egyik résztvevő fehérjére mutatták ki a PS lokalizációt. Mivel a PS fehérjéinek leírása nem teljes és tökéletes, ez az adat felhasználható, hogy olyan új fehérjéket találjunk, amelyekről korábban még nem mutatták ki hogy ebbe a térrészben is előfordulnak: ha megkeressük azokat a fehérjéket, amiknek számos PS interakciós partnerük van, de magáról a fehérjéről ezt még nem mutatták ki, érdekes jelölteket kaphatunk: az Lrrk2 587 PS kapcsolattal rendelkezik, mégsem található meg a négy forrásadatbázis egyikében sem, azonban rövid kutatás után megtalálhatjuk, hogy Lee és munkatársai már bemutatták, hogy az Lrrk2 kölcsönhat az Eif4ebp1 fehérjével az idegvégződés és az izomrost között résben [130]. Más esetben kevésbé direkt bizonyítékok találhatóak: az Ntrk1 szintén nem rendelkezik PS besorolással, annak ellenére hogy 700 PS partnere ismert a PSINDB alapján. Az Ntrk1 egy az idegek növekedésében szerepet játszó fehérje, amely szerepet játszik azok fejlődésben és a differenciálódásban is [131]. Más esetekben a sok PS partnerrel rendelkező fehérjék neurodegeneratív betegségekhez köthetőek: mind az Esr2 [132], mind a Myc [133] kapcsolatba hozhatóak az Alzheimer-kórral.

DOI:10.15774/PPKE.ITK.2023.001



18. 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

5.3.5 A PSINDB jelentősége

A posztgenomikus éra egyik nagy kihívása, hogy a kísérletes technikák egyre nagyobb áteresztőképessége és a technológiák egyre szélesebb körben való elterjedése eredményeként hatalmas mennyiségű biológiai adat keletkezik napról napra. Egyetlen modern szekvenáló készülék naponta több száz terabájtnyi adatot generál [134]. További kihívást jelent, hogy a területen publikált új tudományos eredmények száma exponenciális tendenciát mutat, illetve rendkívül sok adat szétszórtan található meg az irodalomban. Az adatok hozzáférhetősége, rendezettsége és adott esetben értelmezése szempontjából nagyon fontos szerep jut a biológiai adatbázisoknak. A biológiai adatbázisokat sok szempont szerint lehet osztályozni. Léteznek általános és átfogó adatbázisok, mint a UniProt, amely organizmusok lehető legszélesebb gyűjteményét magában foglalja, és több százmillió fehérjéjére ad egységes keretrendszerben leírást. Vannak olyan adatbázisok, amelyek bizonyos organizmusokról (például csak ember, neXtProt), celluláris kompartmenekről (extracelluláris mátrix fehérjék: MatrixDB), vagy szerkezeti csoportról (rendezetlen fehérjék: MobiDB) tárol adatokat. A specifikusan posztszinaptikus adatokat tartalmazó adatbázisok (a doktori témám elindításakor) a SynaptomeDB, G2C és a SynGO adatbázisok voltak (7. táblázat). A SynaptomeDB másodlagos adatbázis (DNS és fehérje szintű információkat is integrál), a G2C a fehérje listák mellett az ismert betegségeket helyezi fókuszba, míg a SynGO a meglévő GO funkcionális és lokalizációs adatokat egészíti ki szinapszis specifikus leírásokkal. Az adatbázisok részletesebb információtartalmát a 5. táblázatban lehet látni.

DOI:10.15774/PPKE.ITK.2023.001

7. táblázat (Poszt)szinapszis specifikus adatbázisok adattartalma és az utolsó frissítés dátuma

Adatbázis	Adatok gyűjtése	Utolsó frissítés
SynaptomeDB	Irodalom kutatás és externális adatbázisok frissítése	2022.08.08.
G2C	Kísérletes: neurológiai preparáció + LC-MS/MS	2011.07.01.
SynGO	Szakértőkből álló konzorcium manuláis kurációja	2021.02.25.

A biológiai adatbázisok egyik szűk keresztmetszete azok karbantartása. Sok esetben a csoport aktuális kutatási iránya vagy finanszírozása miatt létrejövő adatbázis hosszútávú fenntartása nehézzé válik, és gyakran meg is hiúsul. Egy 2015-ben elvégzett, 18 évet átöltelő vizsgálat szerint a megfigyelt több, mint 300 adatbázis ∼60%-a nem volt elérhető [135] a vizsgálat végére, 14%-t pedig archiválták, nem frissítették többé.

A posztszinapszis esetében két, jelenleg már nem elérhető adatbázisról érdemes említést tenni: a SynDB (2007), ami számos funkcionális annotációt foglalt magában a szinaptikus fehérjékre, illetve a SynSysNet (2013), ami SQL adatbázisban gyűjtötte a fehérje-fehérje interakciókat, 3D szerkezetet. A Posztszinaptikus Interakciós (PSINDB) adatbázis felállításakor a saját kutatásunk során szembetűnő limitációkra a fenti két adatbázis legalábbis részleges megoldást nyújtott volna. Az információs korlátok mellett a motivációt a csoport fehérjeszerkezeti fókusza és a posztszinapszis komplexeinek megismerése adta. A PSINDB fejlesztésének megkezdésekor nem volt elérhető olyan adatbázis, ami specifikus a (poszt)szinapszisra, elsődlegesen interakció fókuszú és a PPI-k mellett kiegészítésként tartalmaz más adatokat. 2021 júniusában Sorokina és munkatársai publikálták az adatforrásukat, amely más szemlélet és koncepció szerint született, mint a PSINDB. Az interakciós adatokban valószínűleg jelentős átfedés lehet, azonban míg a PSINDB szerkezeti szinten, funkcionális adatokkal kiegészítve a bináris PPI-ket igyekszik többlet információt adni a posztszinapsziról, addig Sorokina és munkatársai adatbázisukban a fehérjék interakcióit hálózati modellek felállítására használták és elsősorban az egyes neuronális betegségek közötti összefüggésekre világítanak rá az adatokkal (8. táblázat).

DOI:10.15774/PPKE.ITK.2023.001

8. táblázat Specifikus (poszt)szinaptikus adatbázisok összehasonlítása

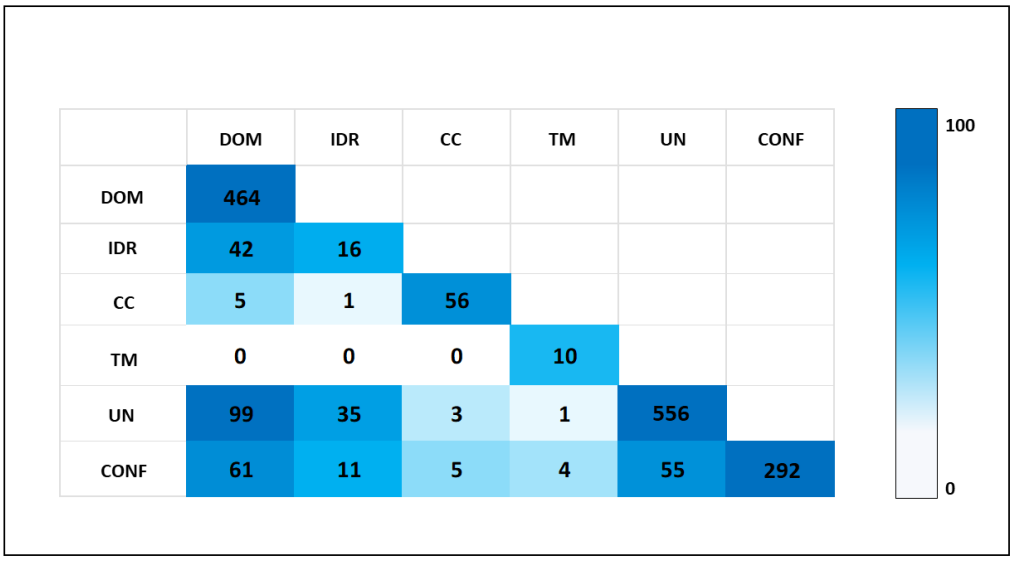
Adatbázis	Sorokina et. al, 2021	Kalman et. al, 2022 (PSINDB)
Fő fókusz	hálózatos megközelítés főleg betegség fókusszal	PPI-ket meghatározó szerkezeti és funkcionális jellemzők
Adattartalom	Teljes szinapszis ∼8000 fehérje	Posztszinapszis ∼2000 fehérje
Fehérje meghatározás	saját gyűjtés publikációkból (30 a posztszinapszisra)	független adatforrások (4 adatbázis)
Interakciós adatok	DIP, IntAct, BioGRID	Manuális kuráció, IntAct, BioGRID, STRING
Kötőrégió	nincs	van
Betegség adatok	OMIM	OMIM, DiseaseOntology
Funkcionális annotáció	GO	GO
Formátum, hozzáférhetőség	SQL	Weboldal + letölthető PSI-MI

5.4 Kötőrégiók elemzése és javaslattétel új régiókra

5.4.1 A kötőrégiók szerkezeti tulajdonságainak elemezése

A PSINDB-ben a kölcsönhatásokon és a kötőrégiókon kívül számos egyéb adat található meg, amelyek a kötésekkel együtt vizsgálva egy komplexebb képet adnak a kölcsönhatásokról. A kötőrégiókat adatgyűjtési és reprezentációs aspektusból is prioritásként kezeltük, ezért az adatbázisban található információk elemzésénél is hangsúlyt fektettünk a vizsgálatokra. Ennek részeként az eltérő részletességű (ld. 5.3.1. fejezet) kötő régiókat elemeztem szerkezeti adatok hozzáadásával és vizsgáltam, hogy milyen egységek között valósul meg interakció. Ahogy a 2.3.1.4. fejezetben is bemutatásra került, a kölcsönhatások jellemzően domének és motívumok révén valósulnak meg, azonban más szerkezeti elemek is részt vehetnek interakciók kialakításában, például a 5.2.7.1. fejezetben említett PIK3R2 coiled-coil régióján keresztül kapcsolódik a partneréhez. Ebben a részben azt vizsgáltam, ha adott egy kötőrégió, és ott meg van határozva milyen szerkezeti elem található (globuláris domén, coiled-coil, transzmembrán vagy rendezetlen aminosavak dominálnak a szakaszon belül), akkor milyen kapcso-

DOI:10.15774/PPKE.ITK.2023.001



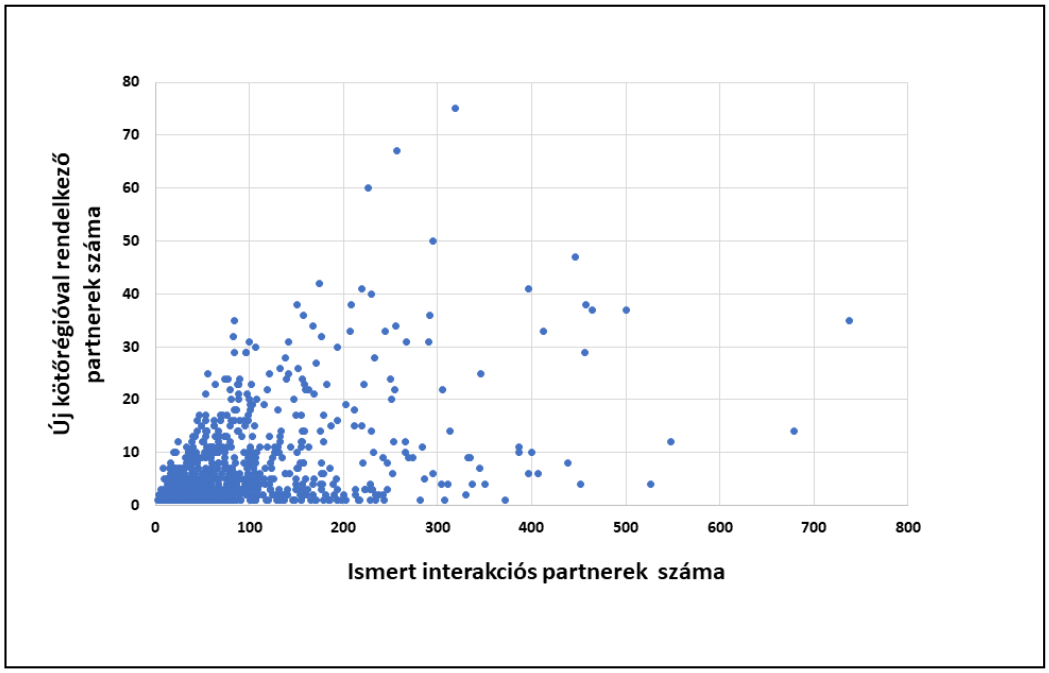
19. ábra A kötőrégiók szerkezeti eloszlása a 'szükséges kötőrégiók' (necessary binding regions) szintjén. A jellemzőbb szerkezeti kapcsolatok - mint a domén-domén és domén-rendezetlen - sötétebb színnel találhatóak meg (dom: domén, idr: rendezetlen, cc: coiled-coil,TM: transmemrbán, un: ismeretlen conf: ellentmondádsos)

lódások jellemzőek. A PSINDB-ben található adatok alapján a 'szükséges kötőrégiók' (necessary binding region) szintjén a PS fehérjékre a legjellemzőbbek a domén-domén, coiled-coil és domén-rendezetlen kölcsönhatások voltak (19. ábra). Az adatok 20%-a esetében ellentmondásos szerekezeti eredményeket kaptunk, ami azt jelenti, hogy nem tudjuk egyértelműen megmondani, hogy milyen elem vesz részt az interakció létrehozásában. Az adatok 37%-ban nem predikált szerkezeti elembe esett a kötőrégió. Megegyező tendenciák látszódnak a 'sufficient binding region' szinten is (ld. Függelék 4. ábra), azonban az ellentmondásos esetek száma növekszik a konkrét szerkezeti eleme közötti kapcsolatok terhére.

5.4.2 Domén-domén interakciók becslése

A meglévő kötőrégiók elemzése mellett fontosnak tartottuk azt is, hogy azokat a nagyobb számban előforduló eseteket is értelmezzük, amikor csupán az interakció tényét ismerjük, azonban nem rendelkezünk információval a kötőrégiókról. Ebből a megközelítésből egy evidens következő lépésnek tűnt, hogy az meglévő bináris párokat használva esetleges kötőrégiókra tegyünk javaslatot. Az 5.3.1. fejezetben ismertetettek szerint, kölcsönhatások kialakulhatnak például globuláris domének és flexibilis régiók segítségével. A posztszinapszis esetében tudjuk, hogy mindkét típusú interakciók fontos szerepet játszanak, de a domén-domén kapcsolatokról sokkal több kísérletes adat áll rendelkezésre.
A vizsgálat során kiszámoltuk a PDB adatbázisban található összes fehérje-fehérje kölcsönhatást (hasonlóan ahogy a PSINDB összeállításakor, ld. 5.3.1. fejezet), majd megnéztük hogy ezek milyen Pfam domének között valósulnak meg. Következő lépésben azt néztük, hogy amennyiben két fehérje

DOI:10.15774/PPKE.ITK.2023.001



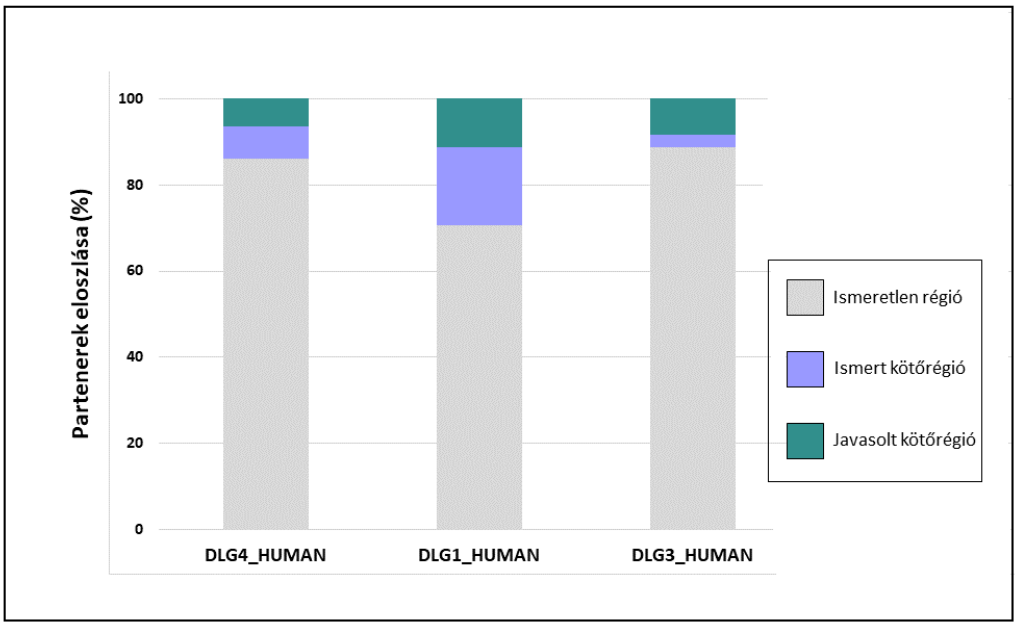
20. ábra A PSINDB fehérjéinek eloszlása - ismert és új kötőrégiók számának alakulása

kölcsönhatásba kerül, rendelkezek-e olyan Pfam domén párral, amely a PDB-ben megjelenik (20. ábra). Az általunk felállított PS adatszettben a 2 160 fehérjéből 845 fehérjére legalább 1 új domének által mediált interakciót határoztunk meg, átlagosan pedig 7 új kötőrégiót mondunk ezekre fehérjékre.

A posztszinaptikus denzitást formáló fehérjék egy fontos családja a MAGUK fehérjék. Ezek fontos alcsaládja a DLG fehérjék, amelyek jellegzetes domén összetétellel rendelkező multidomén fehérjék (ld. Függelék 5. ábra). Ezekre az állvány fehérjékre a PSD "magjaként" tekintünk. A négy fehérjére összesen több száz interakciós partner ismerünk, ahogy részletesebben a 21. ábrán látszik. A DLG4/PSD-95, DLG1/SAP97, illetve a DLG3/SAP102 estében hiába ismerünk számos partnerfehérjét, viszonylag kis hányadban található meg kötőrégió a PSINDB-ben. A négy fehérjéből három esetében domén-domén kapcsolat segítségével tettünk javaslatot a kötőrégióra: a DLG1 esetében 17, a DLG3 esetében 7 és a DLG4 esetében 29 új kötőrégiót azonosítottunk (21. ábra).

A DLG4/PSD-95 és NOS1 közötti kölcsönhatás a két fehérje PDZ doménjei (Pfam: PF000595) között valósulhat meg. Erre nem található meg kötőrégió szintű adat egyik általunk integrált adatbázisban sem, illetve a mi irodalomkutatásunk és gyűjtésünk során sem került elő ez az interakció. Azonban az első cikket, amely ezt az interakciót leírja és kötőrégió szintű információt tartalmaz 1996-ban írták [136]. Biológiailag fontos szerepe van, a két fehérje kapcsolatát (illetve komplexüket az NMDA receptorral) megemlítették már depresszióval összefüggésben [137], illetve iszkémiás agykárosodással kapcsolatban is [138], mindkét esetben külön hangsúlyozva a PDZ domének esetleges szerepét.

DOI:10.15774/PPKE.ITK.2023.001

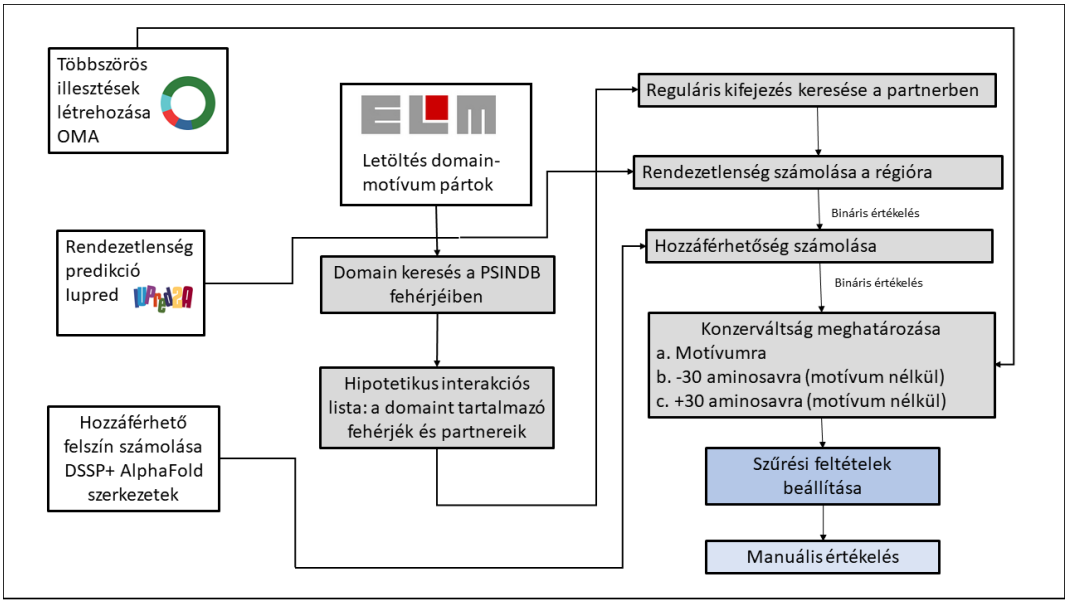


21. ábra DLG család kötőrégióinak megoszlása az ismeretlen, ismert, illetve újonnan javasolt kötőrégiók között. Az ismeretlen régiók szürkével, ismertelek lilával, újonnan javasoltak zölddel jelölve

5.4.3 Motívum-domén interakciók becslése

A domén-domén interakciók mellett az élő szervezetekben legalább két nagyságrenddel több olyan interakció fordul elő, amelyben az egyik partner legalább egy rendezetlen régiót tartalmaz. Ezek közül számos interakció rövid lineáris motívumokon keresztül valósul meg. A posztszinapszis fontos fehérjéiben számos olyan domén megtalálható, amelyről kísérletes adatok bizonyítják, hogy képes rövid lineáris motívumokkal kapcsolatba lépni. A motívum-domén kapcsolatok legmegbízhatóbb forrása az ELM adatbázis (ld. 4.1.6.1. fejezet). Ezek posztszinaptikus tanulmányozására egy olyan pipeline-t állítottam fel, ami a korábbiakkal megegyezően a PSINDB-ben található interakciós adatokat használja fel. A folyamat során a területen használt metódusokat és programokat alkalmaztam úgy, hogy nagyskálás/ félautomatikus legyen az adatok létrehozása/letöltése. Az előkészítő lépésekben a PSINDB-ben meghatározott összes (2160) fehérjére megkerestem az OMA adatbázisból 24 kiválasztott faj ortológ fehérjéit (ld. Függelék 6. ábra). Az ortológokon szekvenciaillesztést hajtottam végre a ClustalOmega programmal,az IuPreddel rendezetlenséget predikáltam a hozzáférhető felszínüket pedig az EMBL-EBI oldaláról letöltött AlphaFold szerkezeteken DSSP-vel határoztam meg (22. ábra, bal). Eközben az ELM adatbázisból letöltöttem a ligandum (LIG) csoporthoz tartozó motívum domén párokat. A doméneket kerestem a PSINDB fehérjéiben és egy hipotetikus listát hoztam létre, amelyben ezek a fehérjék és posztszinaptikus kötőpartnereik szerepeltek. A domént tartalmazó fehérjék partnereiben ezután kerestem a feltételezett motívumokat. Ezeket az adatokat egészítettem ki utána a rendezetlen predikció eredményeivel, ahol a motívum találat régiójában ellenőriztem a kapott IUPred

DOI:10.15774/PPKE.ITK.2023.001



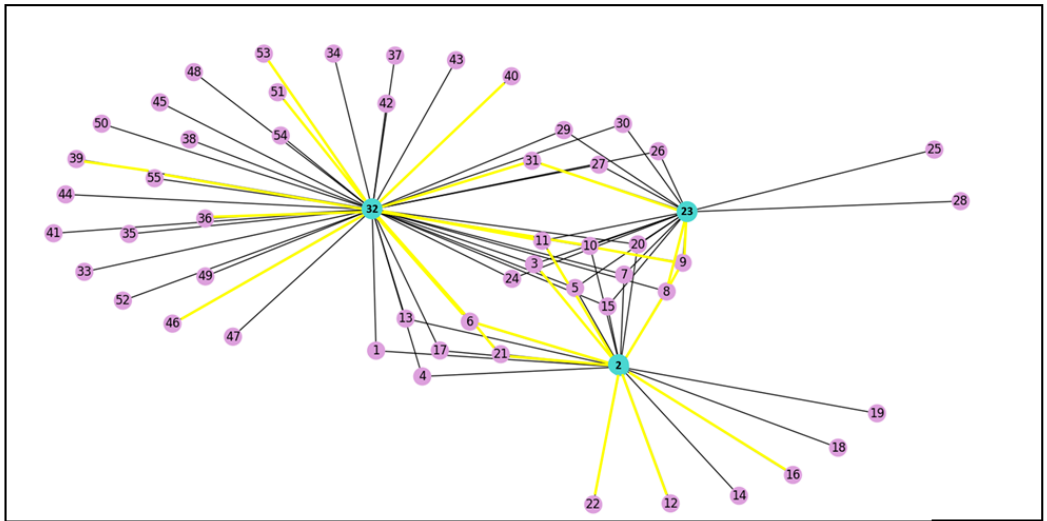
22. ábra A domén-motívum kötőrégiók meghatározásának folyamatábrája

pontszámot és átlag alapján binárisan értékeltem a régiót (átlag 0.5 felett, akkor 1 = rendezetlen). A hozzáférhetőséget is kiszámoltam a régiókra, a módszerek fejezetben korábban leírtak szerint (ld. 4.6. fejezet). A többszörös szekvencia illesztésben a motívumra Shannon-entrópiát számoltam minden egyes pozícióra, majd ezeknek vettem az átlagát a motívumon. A többszörös szekvenciaillesztésben is ellenőriztem a motívumot, és azzal az egyszerűsítéssel éltem, hogy ha gap volt az alignment humán szekvenciára vontakozó részében, azt érvénytelen pontszámmal jelöltem. A folyamat eredményét végül egy MS Excel táblázatba mentettem el. Itt összesen 83 341 találatot kaptam aminek igen jelentős része fals pozitív eredmény volt. Annak érdekében, hogy a hibás találatokat szűrjem, a következő feltételeket alkalmaztam:

- a motívumnak rendezetlen régióba kell esnie
- a motívumnak hozzáférhető felszínen kell lennie
- a Shannon-entalpiának 0.0 kellett lennie

A következő lépés már nem része pipelinenak, de a leszűrt Excel tábla adatait egyszerűen beolvashatóak és hálózatosan ábrázolhatóak a kapott interakciók. Három konkrét posztszinpatikus denzitásban található állvány fehérjén szeretném bemutatni a pipeline működését. Ezek a korábban említett fehérjék a DLG1, DLG3 és a DLG4, amelyekben olyan domének találhatóak, amelyek alkalmasak rövid lineáris motívumok megkötésére (pl. SH3 vagy PDZ) (23. ábra). Abban az esetben, hogyha ezek a fehérjék a domén oldalon találhatóak és a partnerükben megtalálható a lineáris motívum, 152 új kötőrégióra lehet javaslatot tenni (továbbra is alkalmazott kényszerfelételek, 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.

DOI:10.15774/PPKE.ITK.2023.001



1	CTND2_HUMAN	11	CCG2_HUMAN	21	DUS10_HUMAN	31	NLGN3_HUMAN	41	RIMS1_HUMAN	51	SHAN1_HUMAN
2	***DLG1_HUMAN***	12	GPC5B_HUMAN	22	EZRI_HUMAN	*32*	***DLG4_HUMAN***	42	DYN1_HUMAN	52	NRX2A_HUMAN
3	DLGP4_HUMAN	13	IQEC1_HUMAN	*23*	***DLG3_HUMAN***	33	CDKL5_HUMAN	43	FBX2_HUMAN	53	SHAN2_HUMAN
4	KCNJ2_HUMAN	14	LCK_HUMAN	24	CCG4_HUMAN	34	ABL2_HUMAN	44	HSP7C_HUMAN	54	HNRPK_HUMAN
5	TANC2_HUMAN	15	SHSA9_HUMAN	25	NEDD4_HUMAN	35	AT1A1_HUMAN	45	SI1L1_HUMAN	55	ABI1_HUMAN
6	EFNB3_HUMAN	16	EPB41_HUMAN	26	NETO1_HUMAN	36	HGS_HUMAN	46	NLGNX_HUMAN		
7	SHSA7_HUMAN	17	ASAP1_HUMAN	27	SHSA6_HUMAN	37	CADH4_HUMAN	47	SOS1_HUMAN		
8	DLGP1_HUMAN	18	MYO6_HUMAN	28	DC1I1_HUMAN	38	SCN2A_HUMAN	48	ABLM1_HUMAN		
9	GRIK2_HUMAN	19	HCN1_HUMAN	29	ZDHC5_HUMAN	39	KCND2_HUMAN	49	PP1B_HUMAN		
10	CNKR2_HUMAN	20	DLG2_HUMAN	30	KCNA1_HUMAN	40	SYNPO_HUMAN	50	KPCE_HUMAN		

23. ábra 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)

DOI:10.15774/PPKE.ITK.2023.001

5.4.4 A kötőrégiók vizsgálatának eredményei

A fehérje interakciós adatok rendszerezése során a legtöbb adatbázis nem rögzíti a kötőrégiókat. Ennek több magyarázata is lehet, többek között az, hogy a kötőrégiók biztos megállapítása rendkívül időigényes is lehet. A posztszinapszis komplexeinek tényleges megismeréséhez azonban szükséges lépés a kötések molekuláris szintű feltárása, amelynek egy kezdeni lépése lehet a bináris interakciók részletes megismerése. A domén-domén interakciók megbízhatóságának vizsgálatára a kapott eredményt összevetettem egy tavaly publikált módszerrel a PPIDomainMiner eredményeivel (ld. Függelék 7. ábra), akik hasonló adatokkal dolgoznak, azonban más megközelítéssel. Ezen összehasonlítás alapján az látszik, hogy a mostani módszer valószínűleg túl megengedő, azonban a DLG4/nNOS példa alapján rávilágíthat olyan összefüggésekre amiket korábban nem vagy csak részben vizsgáltak, így segítve az annotációs munkát. A posztszinapszis fehérjéi között számos domén-domén kölcsönhatás már jól tanulmányozott, és sok ezekből szerepel is megfelelő adatbázisokban, azonban a tranziens kölcsönhatásokról, amelyeknek jellemző előfordulása a mi vizsgálatunkból is látszik, még nem igazán látunk átfogó képet. Bár létezik több motívum predikciós eljárás is, én egy új rendszert állítottam fel, amelynek két oka volt:

1. alapvetően webszerverként működnek, ezért nagy mennyiségű fehérje esetében nehezen használhatóak pl.Slimsearch, ELM predikciós része.

2. ezek az eljárások legtöbbször nem veszik figyelembe a már ismert fehérje-fehérje kölcsönhatási adatokat, amik viszont nekem a rendelkezésre álltak.

Az első megközelítéssel kapott adatok alapján azonban jól látszik, hogy a posztszinapszis esetében is fontos lenne a tranziens kölcsönhatások széleskörű feltérképezése.

DOI:10.15774/PPKE.ITK.2023.001

6 Összefoglalás és kitekintés

A posztszinapszis (PS) a neuronális jelátvitel fundamentális alegysége. Az elmúlt több, mint két évtizedben rengeteg kísérletet végeztek működésének feltárására, azonban ismereteink az organellumról még mindig részlegesek: a posztszinapszis pontos fehérje összetétele nem ismert, és emellett számos az alapvető működésben szerepet játszó folyamatot még csak részben értünk - elég csak a fázisszeparáció jelenségére gondolni, amiről az első átfogó leírások csak az elmúlt pár évben jelentek meg. Bár az *in silico* módszerek számos limitációval rendelkeznek a kísérletes technikákkal szemben, segítségükkel dedikált adatbázisok és nagyskálás elemzések hozhatóak létre. A doktori munkám során számítógépes eljárásokkal többféle aspektusból, rendszerszinten vizsgáltam a posztszinapszis fehérjéit és létrehoztam átfogó gyűjteményüket (PostSynapticInteractionDataBase, PSINDB).

A posztszinapszis sokrétű működéséért a sejt dinamikusan változó fehérjehálózata felelős. A körülbelül kétezer fehérjét tartalmazó hálózatban akár több százezer különböző kölcsönhatás is kialakulhat. A kapcsolatok kialakítását meghatározza a fehérjék szerkezete, ezek az elemek (más szóval modulok) önmagukban, illetve különböző kombinációkban felelősek az interakciók kialakításáért. Ismereteink szerint a PS esetében kiemelt jelentősége van a globuláris domén-domén interakcióknak, valamint motívumok által vezérelt kölcsönhatásoknak. Eredményeim azonban arra is egyértelműen rávilágítottak, hogy az ilyen szempontból eddig kevésbé vizsgált szerkezeti elemeken (mint a coiled-coil) keresztül megvalósuló interakciók is jelentősen befolyásolhatják a PS működését. A coiled-coil fehérjék magas érintettsége a központi idegrendszeri betegségekben kiemeli ezen multimerizációs motívum által (is) összetartott fehérjekomplexek jelentőségét. Az egyes komponens fehérjék által kialakítható kölcsönhatások rendszerezése többek között megnyitja az utat olyan rendszerbiológiai modellek felállítása előtt, melyek explicit módon képesek figyelembe venni a fehérje komplexek összetételét és annak különböző hatásokra - expressziós szint, poszttranszlációs módosítások, mutációk stb. - történő megváltozását, hozzájárulva ezek funkcionális jelentőségének megértéséhez [129].

A posztszinapszis fehérjéihez több száz betegség kötődik, melyek pontos molekuláris mechanizmusai még feltáratlanok. Ezek közül számos több millió embert érint szerte a világon, ugyanakkor a pontos működésük megértése nélkül azok kezelése is jelentős limitációkba ütközik, ahogyan azt például az Alzheimer-kór esetében látjuk is. A PS működésének minél részletesebb megértésével nem csak neuronális betegségek gyógyítására nyílna lehetőség, hanem gondolkodásunk alapjaira is fény derülhetne, amellyel megválaszolnánk az emberiség története óta fennálló egyik legfontosabb biológiai kérdést, nevezetesen, hogy miképp is működik agyunk.

DOI:10.15774/PPKE.ITK.2023.001

7 Köszönetnyilvánítás

Először is hatalmas köszönettel tartozom témavezetőmnek Dr. Gáspári Zoltánnak, aki engedte, hogy megtapasztaljam a kutatás szabadságát, ugyanakkor mindig visszaterelt amikor már túl messzire kalandoztam. A szakmai és emberi példamutatása remélem iránytűként fog szolgálni egész életemben.
Köszönettel tartozom az Egyetem vezetésének, elsősorban Dr. Iván Kristófnak, hogy doktori tanulmányaim során a szükséges feltételeken túl is számos lehetőséget biztosított számunkra, továbbá egy olyan kiemelkedő szakmai közeget, ahol mi doktoranduszok a munkán túl is mindig jól érezhettük magunkat.
Szeretném hálámat kifejezni azoknak a kutatóknak és mentoraimnak, akik a bioinformatikai pálya felé tereltek, illetve az utamat egyengették, Dr. Búzás Zsuzsannának, Prof. Dr. Pongor Sándornak, Prof. Dr. Orosz Lászlónak, Dr. Barta Endrének és Dr. Stéger Viktornak.
Hálával tartozom Prof. Toby Gibsonnak, akitől a rövid idő ellenére is elképesztően sokat tanulhattam és akinek köszönhetően egy teljesen új biológiai szemléletet nyertem.
Szeretném megköszönni Dr. Dudola Dánielnek a PSINDB-vel kapcsolatos minden munkáját, nélküle biztosan nem jöhetett volna létre az adatbázis. Hatalmas külön köszönet illeti Dr. Fichó Erzsébetet is, aki rengeteget segített a dolgozatom legutolsó szakaszában és akinek az értékes megjegyzéseiből a doktori tanulmányom legutolsó pillanatáig tanulni tudtam.
Nagy szeretettel és hálával gondolok doktorandusz társaimra, elsősorban Nagy-Kanta Eszterre és Harmat Zitára, akiknek köszönhetően az első években minden bajtársi támogatást megkaptam. Illetve szeretném megköszönni Jesus Alvarado Valverdének minden kedvességét a kinntlétem során.
A doktori munkám során nem csak elmélyíthettem tudásom a fehérjék világában, hanem egy életre szóló társam is lett. Köszönetet szeretnék mondani Dr. Dobson Lászlónak, aki már a kezdettől is maximálisan támogatott szakmailag és akinél mind munkatársként, mind férjként nem tudok elképzelni jobb embert.
Szeretnék köszönetet mondani szüleimnek, akik végtelen türelemmel és támogatással kísérték az elmúlt négy évet. Végül, de nem utolsó sorban szeretném megköszönni a barátaimnak, akikre mindig számíthattam.

DOI:10.15774/PPKE.ITK.2023.001

8 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 (CerEla1. 0) to develop X-and Y-chromosome-linked STR markers. PLoS One, 15(11), e0242506.

Nem referált folyóirat

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

DOI:10.15774/PPKE.ITK.2023.001

9 Irodalomjegyzék

References

[1] Zhe Feng, Xudong Chen, Menglong Zeng, and Mingjie Zhang. Phase separation as a mechanism for assembling dynamic postsynaptic density signalling complexes. *Current opinion in neurobiology*, 57:1–8, 2019.

[2] Seth GN Grant. Synapse diversity and synaptome architecture in human genetic disorders. *Human Molecular Genetics*, 28(R2):R219–R225, 2019.

[3] Malgorzata Borczyk, Kasia Radwanska, and K Peter Giese. The importance of ultrastructural analysis of memory. *Brain Research Bulletin*, 173:28–36, 2021.

[4] Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven Siegelbaum, A James Hudspeth, Sarah Mack, et al. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

[5] Seth GN Grant. Synapse molecular complexity and the plasticity behaviour problem. *Brain and Neuroscience Advances*, 2:2398212818810685, 2018.

[6] Morgan Sheng and Eunjoon Kim. The postsynaptic organization of synapses. *Cold Spring Harbor perspectives in biology*, 3(12):a005678, 2011.

[7] Marcelo P Coba, Andrew J Pocklington, Mark O Collins, Maksym V Kopanitsa, Rachel T Uren, Sajani Swamy, Mike DR Croning, Jyoti S Choudhary, and Seth GN Grant. Neurotransmitters drive combinatorial multistate postsynaptic density networks. *Science signaling*, 2(68):ra19–ra19, 2009.

[8] Frank Koopmans, Pim van Nierop, Maria Andres-Alonso, Andrea Byrnes, Tony Cijsouw, Marcelo P Coba, L Niels Cornelisse, Ryan J Farrell, Hana L Goldschmidt, Daniel P Howrigan, et al. Syngo: an evidence-based, expert-curated knowledge base for the synapse. *Neuron*, 103(2):217–234, 2019.

[9] Matthew J Broadhead, Mathew H Horrocks, Fei Zhu, Leila Muresan, Ruth Benavides-Piccione, Javier DeFelipe, David Fricker, Maksym V Kopanitsa, Rory R Duncan, David Klenerman, et al. Psd95 nanoclusters are postsynaptic building blocks in hippocampus circuits. *Scientific reports*, 6(1):1–14, 2016.

[10] Takeshi Kaizuka and Toru Takumi. Postsynaptic density proteins and their involvement in neurodevelopmental disorders. *The Journal of Biochemistry*, 163(6):447–455, 2018.

[11] Ayse Dosemeci, Richard J Weinberg, Thomas S Reese, and Jung-Hwa Tao-Cheng. The post-synaptic density: there is more than meets the eye. *Frontiers in synaptic neuroscience*, 8:23, 2016.

[12] Mélissa Cizeron, Zhen Qiu, Babis Koniaris, Ragini Gokhale, Noboru H Komiyama, Erik Fransén, and Seth GN Grant. A brainwide atlas of synapses across the mouse life span. *Science*, 369(6501):270–275, 2020.

[13] Guanhua Bai and Mingjie Zhang. Mesophasic assembly of inhibitory postsynaptic density. *Neuroscience Bulletin*, 37(1):141–143, 2021.

[14] Brent Wilkinson and Marcelo P Coba. Molecular architecture of postsynaptic interactomes. *Cellular Signalling*, 76:109782, 2020.

[15] Anders Liljas, Lars Liljas, Goran Lindblom, Poul Nissen, Morten Kjeldgaard, and Miriam-rose Ash. *Textbook of structural biology*, volume 8. World Scientific, 2016.

[16] A Leninger, D Nelson, and M Cox. Lehninger principles of biochemistry, 2017.

[17] Christian B Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 1973.

[18] Cyrus Levinthal. Are there pathways for protein folding? *Journal de chimie physique*, 65:44–45, 1968.

[19] Song-Ho Chong and Sihyun Ham. Folding free energy landscape of ordered and intrinsically disordered proteins. *Scientific reports*, 9(1):1–9, 2019.

[20] S Walter Englander and Leland Mayne. The nature of protein folding pathways. *Proceedings of the National Academy of Sciences*, 111(45):15873–15880, 2014.

[21] Julie S Valastyan and Susan Lindquist. Mechanisms of protein-folding diseases at a glance. *Disease models & mechanisms*, 7(1):9–14, 2014.

[22] Andrei N Lupas, Joana Pereira, Vikram Alva, Felipe Merino, Murray Coles, and Marcus D Hartmann. The breakthrough in protein structure prediction. *Biochemical journal*, 478(10):1885–1890, 2021.

[23] Derek N Woolfson. The design of coiled-coil structures and assemblies. *Advances in protein chemistry*, 70:79–112, 2005.

[24] Jody M Mason and Katja M Arndt. Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem*, 5(2):170–176, 2004.

[25] Linda Truebestein and Thomas A Leonard. Coiled-coils: The long and short of it. *Bioessays*, 38(9):903–916, 2016.

[26] 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.

[27] Gevorg Grigoryan and Amy E Keating. Structural specificity in coiled-coil interactions. *Current opinion in structural biology*, 18(4):477–483, 2008.

[28] 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.

[29] Andrei N Lupas and Markus Gruber. The structure of α-helical coiled coils. *Advances in protein chemistry*, 70:37–38, 2005.

[30] Matthew R Hicks, David V Holberton, Christopher Kowalczyk, and Derek N Woolfson. Coiled-coil assembly by peptides with non-heptad sequence motifs. *Folding and Design*, 2(3):149–158, 1997.

[31] Asmit Bhowmick, David H Brookes, Shane R Yost, H Jane Dyson, Julie D Forman-Kay, Daniel Gunter, Martin Head-Gordon, Gregory L Hura, Vijay S Pande, David E Wemmer, et al. Finding our way in the dark proteome. *Journal of the American Chemical Society*, 138(31):9730–9742, 2016.

[32] Antonio Deiana, Sergio Forcelloni, Alessandro Porrello, and Andrea Giansanti. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PloS one*, 14(8):e0217889, 2019.

[33] Robin Van Der Lee, Marija Buljan, Benjamin Lang, Robert J Weatheritt, Gary W Daughdrill, A Keith Dunker, Monika Fuxreiter, Julian Gough, Joerg Gsponer, David T Jones, et al. Classification of intrinsically disordered regions and proteins. *Chemical reviews*, 114(13):6589–6631, 2014.

[34] Peter Tompa. Intrinsically unstructured proteins. *Trends in biochemical sciences*, 27(10):527–533, 2002.

[35] Norman E Davey, Kim Van Roey, Robert J Weatheritt, Grischa Toedt, Bora Uyar, Brigitte Altenberg, Aidan Budd, Francesca Diella, Holger Dinkel, and Toby J Gibson. Attributes of short linear motifs. *Molecular BioSystems*, 8(1):268–281, 2012.

[36] Christine A Orengo, Annabel E Todd, and Janet M Thornton. From protein structure to function. *Current opinion in structural biology*, 9(3):374–382, 1999.

[37] Nelson Perdigão, Agostinho C Rosa, and Seán I O'Donoghue. The dark proteome database. *BioData mining*, 10(1):1–11, 2017.

[38] James W Fairman, Nicholas Noinaj, and Susan K Buchanan. The structural biology of β-barrel membrane proteins: a summary of recent reports. *Current opinion in structural biology*, 21(4):523–531, 2011.

[39] Lisa N Kinch and Nick V Grishin. Evolution of protein structures and functions. *Current opinion in structural biology*, 12(3):400–408, 2002.

[40] Ian R Humphreys, Jimin Pei, Minkyung Baek, Aditya Krishnakumar, Ivan Anishchenko, Sergey Ovchinnikov, Jing Zhang, Travis J Ness, Sudeep Banjade, Saket R Bagde, et al. Computed structures of core eukaryotic protein complexes. *Science*, 374(6573):eabm4805, 2021.

[41] Ole N Jensen. Interpreting the protein language using proteomics. *Nature reviews Molecular cell biology*, 7(6):391–403, 2006.

[42] Guangyou Duan and Dirk Walther. The roles of post-translational modifications in the context of protein interaction networks. *PLoS computational biology*, 11(2):e1004049, 2015.

[43] Adam P Lothrop, Matthew P Torres, and Stephen M Fuchs. Deciphering post-translational modification codes. *FEBS letters*, 587(8):1247–1257, 2013.

[44] Bruce T Seet, Ivan Dikic, Ming-Ming Zhou, and Tony Pawson. Reading protein modifications with interaction domains. *Nature reviews Molecular cell biology*, 7(7):473–483, 2006.

[45] 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.

[46] Xuan Yang and Andrey A Ivanov. Computational structural modeling to discover ppi modulators. In *Protein–Protein Interaction Regulators*, pages 87–108. 2020.

[47] Dana Reichmann, Ofer Rahat, Shira Albeck, Ran Meged, Orly Dym, and Gideon Schreiber. The modular architecture of protein–protein binding interfaces. *Proceedings of the National Academy of Sciences*, 102(1):57–62, 2005.

[48] 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.

[49] Erica A Golemis, Erica Golemis, and Peter David Adams. *Protein-protein interactions: a molecular cloning manual*. CSHL Press, 2005.

[50] 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.

[51] Kenji Sugase, H Jane Dyson, and Peter E Wright. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 447(7147):1021–1025, 2007.

[52] Javier De Las Rivas and Celia Fontanillo. Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS computational biology*, 6(6):e1000807, 2010.

[53] Till Siebenmorgen and Martin Zacharias. Computational prediction of protein–protein binding affinities. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(3):e1448, 2020.

[54] Martin Zacharias. Accounting for conformational changes during protein–protein docking. *Current opinion in structural biology*, 20(2):180–186, 2010.

[55] Menglong Zeng, Yuan Shang, Yoichi Araki, Tingfeng Guo, Richard L Huganir, and Mingjie Zhang. Phase transition in postsynaptic densities underlies formation of synaptic complexes and synaptic plasticity. *Cell*, 166(5):1163–1175, 2016.

[56] Jinwei Zhu, Qingqing Zhou, Yuan Shang, Hao Li, Mengjuan Peng, Xiao Ke, Zhuangfeng Weng, Rongguang Zhang, Xuhui Huang, Shawn SC Li, et al. Synaptic targeting and function of sapaps mediated by phosphorylation-dependent binding to psd-95 maguks. *Cell reports*, 21(13):3781–3793, 2017.

[57] C Geoffrey Lau and R Suzanne Zukin. Nmda receptor trafficking in synaptic plasticity and neuropsychiatric disorders. *Nature Reviews Neuroscience*, 8(6):413–426, 2007.

[58] Zeynep Öztürk, Cahir J O'Kane, and Juan José Pérez-Moreno. Axonal endoplasmic reticulum dynamics and its roles in neurodegeneration. *Frontiers in neuroscience*, 14:48, 2020.

[59] Akhilesh Kumar Bajpai, Sravanthi Davuluri, Kriti Tiwary, Sithalechumi Narayanan, Sailaja Oguru, Kavyashree Basavaraju, Deena Dayalan, Kavitha Thirumurugan, and Kshitish K Acharya. Systematic comparison of the protein-protein interaction databases from a user's perspective. *Journal of Biomedical Informatics*, 103:103380, 2020.

[60] Eleanor E Deschner, Julia S Lytle, George Wong, Jeanne F Ruperto, and Harold L Newmark. The effect of dietary omega-3 fatty acids (fish oil) on azoxymethanol-induced focal areas of dysplasia and colon tumor incidence. *Cancer*, 66(11):2350–2356, 1990.

[61] Sandra Orchard, Lukasz Salwinski, Samuel Kerrien, Luisa Montecchi-Palazzi, Matthias Oesterheld, Volker Stümpflen, Arnaud Ceol, Andrew Chatr-Aryamontri, John Armstrong, Peter Woollard, et al. The minimum information required for reporting a molecular interaction experiment (mimix). *Nature biotechnology*, 25(8):894–898, 2007.

[62] Bruno Aranda, P Achuthan, Yasmin Alam-Faruque, I Armean, Alan Bridge, C Derow, Marc Feuermann, AT Ghanbarian, Samuel Kerrien, Jyoti Khadake, et al. The intact molecular interaction database in 2010. *Nucleic acids research*, 38(suppl_1):D525–D531, 2010.

[63] 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.

[64] Catarina D Campbell and Evan E Eichler. Properties and rates of germline mutations in humans. *Trends in Genetics*, 29(10):575–584, 2013.

[65] 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.

[66] Satoshi Oota. Somatic mutations–evolution within the individual. *Methods*, 176:91–98, 2020.

[67] Yao-Fu Chang, J Saadi Imam, Miles F Wilkinson, et al. The nonsense-mediated decay rna surveillance pathway. *Annual review of biochemistry*, 76(1):51–74, 2007.

[68] Pedro Morais, Hironori Adachi, and Yi-Tao Yu. Suppression of nonsense mutations by new emerging technologies. *International Journal of Molecular Sciences*, 21(12):4394, 2020.

[69] Zhen Wang and John Moult. Snps, protein structure, and disease. *Human mutation*, 17(4):263–270, 2001.

[70] Mu Gao, Hongyi Zhou, and Jeffrey Skolnick. Insights into disease-associated mutations in the human proteome through protein structural analysis. *Structure*, 23(7):1362–1369, 2015.

[71] Robert E Steward, Malcolm W MacArthur, Roman A Laskowski, and Janet M Thornton. Molecular basis of inherited diseases: a structural perspective. *TRENDS in Genetics*, 19(9):505–513, 2003.

[72] Poh Hui Chia, Franklin Lei Zhong, Shinsuke Niwa, Carine Bonnard, Kagistia Hana Utami, Ruizhu Zeng, Hane Lee, Ascia Eskin, Stanley F Nelson, William H Xie, et al. A homozygous loss-of-function camk2a mutation causes growth delay, frequent seizures and severe intellectual disability. *Elife*, 7, 2018.

[73] Jason Gandhi, Anthony C Antonelli, Adil Afridi, Sohrab Vatsia, Gunjan Joshi, Victor Romanov, Ian VJ Murray, and Sardar Ali Khan. Protein misfolding and aggregation in neurodegenerative

diseases: a review of pathogeneses, novel detection strategies, and potential therapeutics. *Reviews in the Neurosciences*, 30(4):339–358, 2019.

[74] Claudio Soto and Sandra Pritzkow. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nature neuroscience*, 21(10):1332–1340, 2018.

[75] Roshni Bhattacharya, Peter W Rose, Stephen K Burley, and Andreas Prlić. Impact of genetic variation on three dimensional structure and function of proteins. *PloS one*, 12(3):e0171355, 2017.

[76] Yum L Yip, Maria Famiglietti, Arnaud Gos, Paula D Duek, Fabrice PA David, Alain Gateau, and Amos Bairoch. Annotating single amino acid polymorphisms in the uniprot/swiss-prot knowledgebase. *Human mutation*, 29(3):361–366, 2008.

[77] UniProt Consortium. Uniprot: a worldwide hub of protein knowledge. *Nucleic acids research*, 47(D1):D506–D515, 2019.

[78] Mehdi Pirooznia, Tao Wang, Dimitrios Avramopoulos, David Valle, Gareth Thomas, Richard L Huganir, Fernando S Goes, James B Potash, and Peter P Zandi. Synaptomedb: an ontology-based knowledgebase for synaptic genes. *Bioinformatics*, 28(6):897–899, 2012.

[79] Àlex Bayés, Mark O Collins, Mike DR Croning, LOUIE N Van De Lagemaat, Jyoti S Choudhary, and Seth GN Grant. Comparative study of human and mouse postsynaptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. 2012.

[80] Robert D Finn, Jody Clements, and Sean R Eddy. Hmmer web server: interactive sequence similarity searching. *Nucleic acids research*, 39(suppl_2):W29–W37, 2011.

[81] Ian Sillitoe, Tony E Lewis, Alison Cuff, Sayoni Das, Paul Ashford, Natalie L Dawson, Nicholas Furnham, Roman A Laskowski, David Lee, Jonathan G Lees, et al. Cath: comprehensive structural and functional annotations for genome sequences. *Nucleic acids research*, 43(D1):D376–D381, 2015.

[82] Alexey G Murzin, Steven E Brenner, Tim Hubbard, and Cyrus Chothia. Scop: a structural classification of proteins database for the investigation of sequences and structures. *Journal of molecular biology*, 247(4):536–540, 1995.

[83] Sara El-Gebali, Jaina Mistry, Alex Bateman, Sean R Eddy, Aurélien Luciani, Simon C Potter, Matloob Qureshi, Lorna J Richardson, Gustavo A Salazar, Alfredo Smart, et al. The pfam protein families database in 2019. *Nucleic acids research*, 47(D1):D427–D432, 2019.

[84] Stephen K Burley, Helen M Berman, Gerard J Kleywegt, John L Markley, Haruki Nakamura, and Sameer Velankar. Protein data bank (pdb): the single global macromolecular structure archive. *Protein Crystallography*, pages 627–641, 2017.

[85] Rose Oughtred, Chris Stark, Bobby-Joe Breitkreutz, Jennifer Rust, Lorrie Boucher, Christie Chang, Nadine Kolas, Lara O'Donnell, Genie Leung, Rochelle McAdam, et al. The biogrid interaction database: 2019 update. *Nucleic acids research*, 47(D1):D529–D541, 2019.

[86] Damian Szklarczyk, Annika L Gable, Katerina C Nastou, David Lyon, Rebecca Kirsch, Sampo Pyysalo, Nadezhda T Doncheva, Marc Legeay, Tao Fang, Peer Bork, et al. The string database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research*, 49(D1):D605–D612, 2021.

[87] Manjeet Kumar, Marc Gouw, Sushama Michael, Hugo Sámano-Sánchez, Rita Pancsa, Juliana Glavina, Athina Diakogianni, Jesús Alvarado Valverde, Dayana Bukirova, Jelena Čalyševa, et al. Elm—the eukaryotic linear motif resource in 2020. *Nucleic acids research*, 48(D1):D296–D306, 2020.

[88] Bálint Mészáros, Gábor Erdős, Beáta Szabó, Éva Schád, Ágnes Tantos, Rawan Abukhairan, Tamás Horváth, Nikoletta Murvai, Orsolya P Kovács, Márton Kovács, et al. Phasepro: the database of proteins driving liquid–liquid phase separation. *Nucleic acids research*, 48(D1):D360–D367, 2020.

[89] László Dobson, István Reményi, and Gábor E Tusnády. The human transmembrane proteome. *Biology direct*, 10(1):1–18, 2015.

[90] Adrian M Altenhoff, Clément-Marie Train, Kimberly J Gilbert, Ishita Mediratta, Tarcisio Mendes de Farias, David Moi, Yannis Nevers, Hale-Seda Radoykova, Victor Rossier, Alex Warwick Vesztrocy, et al. Oma orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic acids research*, 49(D1):D373–D379, 2021.

[91] Zoltán Gáspári, Dániel Süveges, András Perczel, László Nyitray, and Gábor Tóth. Charged single alpha-helices in proteomes revealed by a consensus prediction approach. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1824(4):637–646, 2012.

[92] Ákos Kovács, Dániel Dudola, László Nyitray, Gábor Tóth, Zoltán Nagy, and Zoltán Gáspári. Detection of single alpha-helices in large protein sequence sets using hardware acceleration. *Journal of Structural Biology*, 204(1):109–116, 2018.

[93] Stephen F Altschul, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, 1990.

[94] Fabian Sievers and Desmond G Higgins. Clustal omega. *Current protocols in bioinformatics*, 48(1):3–13, 2014.

[95] Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. Cd-hit: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152, 2012.

[96] Gene Ontology Consortium. Gene ontology consortium: going forward. *Nucleic acids research*, 43(D1):D1049–D1056, 2015.

[97] Jan Ludwiczak, Aleksander Winski, Krzysztof Szczepaniak, Vikram Alva, and Stanislaw Dunin-Horkawicz. Deepcoil—a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics*, 35(16):2790–2795, 2019.

[98] Andrew V McDonnell, Taijiao Jiang, Amy E Keating, and Bonnie Berger. Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics*, 22(3):356–358, 2006.

[99] Mauro Delorenzi and Terry Speed. An hmm model for coiled-coil domains and a comparison with pssm-based predictions. *Bioinformatics*, 18(4):617–625, 2002.

[100] Andrei Lupas. Coiled coils: new structures and new functions. *Trends in biochemical sciences*, 21(10):375–382, 1996.

[101] Thomas L Vincent, Peter J Green, and Derek N Woolfson. Logicoil—multi-state prediction of coiled-coil oligomeric state. *Bioinformatics*, 29(1):69–76, 2013.

[102] John Walshaw and Derek N Woolfson. Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *Journal of molecular biology*, 307(5):1427–1450, 2001.

[103] Joost Schymkowitz, Jesper Borg, Francois Stricher, Robby Nys, Frederic Rousseau, and Luis Serrano. The foldx web server: an online force field. *Nucleic acids research*, 33(suppl_2):W382–W388, 2005.

[104] Bálint Mészáros, Gábor Erdős, and Zsuzsanna Dosztányi. Iupred2a: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic acids research*, 46(W1):W329–W337, 2018.

[105] Lynn M Schriml, Elvira Mitraka, James Munro, Becky Tauber, Mike Schor, Lance Nickle, Victor Felix, Linda Jeng, Cynthia Bearer, Richard Lichenstein, et al. Human disease ontology 2018 update: classification, content and workflow expansion. *Nucleic acids research*, 47(D1):D955–D962, 2019.

[106] James B Procter, G Carstairs, Ben Soares, Kira Mourão, T Charles Ofoegbu, Daniel Barton, Lauren Lui, Anne Menard, Natasha Sherstnev, David Roldan-Martinez, et al. Alignment of

biological sequences with jalview. In *Multiple Sequence Alignment*, pages 203–224. Springer, 2021.

[107] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12):2577–2637, 1983.

[108] Evgeny Krissinel. Crystal contacts as nature's docking solutions. *Journal of computational chemistry*, 31(1):133–143, 2010.

[109] Thomas D Goddard, Conrad C Huang, Elaine C Meng, Eric F Pettersen, Gregory S Couch, John H Morris, and Thomas E Ferrin. Ucsf chimerax: Meeting modern challenges in visualization and analysis. *Protein Science*, 27(1):14–25, 2018.

[110] Matthew Z Tien, Austin G Meyer, Dariya K Sydykova, Stephanie J Spielman, and Claus O Wilke. Maximum allowed solvent accessibilites of residues in proteins. *PloS one*, 8(11):e80635, 2013.

[111] László Dobson, Bálint Mészáros, and Gábor E Tusnády. Structural principles governing disease-causing germline mutations. *Journal of molecular biology*, 430(24):4955–4970, 2018.

[112] Kaavya A Mohanasundaram, Mani P Grover, Tamsyn M Crowley, Andrzej Goscinski, and Merridee A Wouters. Mapping genotype–phenotype associations of nssnps in coiled-coil oligomerization domains of the human proteome. *Human mutation*, 38(10):1378–1393, 2017.

[113] Peter Y Chou and Gerald D Fasman. Prediction of protein conformation. *Biochemistry*, 13(2):222–245, 1974.

[114] Dániel Süveges, Zoltán Gáspári, Gábor Tóth, and László Nyitray. Charged single α-helix: A versatile protein structural motif. *Proteins: Structure, Function, and Bioinformatics*, 74(4):905–916, 2009.

[115] Romain A Studer, Pascal-Antoine Christin, Mark A Williams, and Christine A Orengo. Stability-activity tradeoffs constrain the adaptive evolution of rubisco. *Proceedings of the National Academy of Sciences*, 111(6):2223–2228, 2014.

[116] Lev G Goldfarb, Marinos C Dalakas, et al. Tragedy in a heartbeat: malfunctioning desmin causes skeletal and cardiac muscle disease. *The Journal of clinical investigation*, 119(7):1806–1813, 2009.

[117] Reka P Toth and Julie D Atkin. Dysfunction of optineurin in amyotrophic lateral sclerosis and glaucoma. *Frontiers in immunology*, 9:1017, 2018.

[118] Gaetano Terrone, Norine Voisin, Ali Abdullah Alfaiz, Gerarda Cappuccio, Giuseppina Vitiello, Nicolas Guex, Alessandra D'Amico, A James Barkovich, Nicola Brunetti-Pierri, Ennio Del Giudice, et al. De novo pik3r2 variant causes polymicrogyria, corpus callosum hyperplasia and focal cortical dysplasia. *European Journal of Human Genetics*, 24(9):1359–1362, 2016.

[119] Balázs Szappanos, Dániel Süveges, László Nyitray, András Perczel, and Zoltán Gáspári. Folded-unfolded cross-predictions and protein evolution: the case study of coiled-coils. *FEBS letters*, 584(8):1623–1627, 2010.

[120] 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.

[121] 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.

[122] 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.

[123] 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.

[124] Petri Kursula. Shanks—multidomain molecular scaffolds of the postsynaptic density. *Current Opinion in Structural Biology*, 54:122–128, 2019.

[125] Marisa K Baron, Tobias M Boeckers, Bianca Vaida, Salem Faham, Mari Gingery, Michael R Sawaya, Danielle Salyer, Eckart D Gundelfinger, and James U Bowie. An architectural framework that may lie at the core of the postsynaptic density. *Science*, 311(5760):531–535, 2006.

[126] Mariko Kato Hayashi, Chunyan Tang, Chiara Verpelli, Radhakrishnan Narayanan, Marissa H Stearns, Rui-Ming Xu, Huilin Li, Carlo Sala, and Yasunori Hayashi. The postsynaptic density proteins homer and shank form a polymeric network structure. *cell*, 137(1):159–171, 2009.

[127] Carlo Sala, Cinzia Vicidomini, Ilaria Bigi, Adele Mossa, and Chiara Verpelli. Shank synaptic scaffold proteins: keys to understanding the pathogenesis of autism and other synaptic disorders. *Journal of neurochemistry*, 135(5):849–858, 2015.

[128] Michael Bucher, Stephan Niebling, Yuhao Han, Dmitry Molodenskiy, Fatemeh Hassani Nia, Hans-Jürgen Kreienkamp, Dmitri Svergun, Eunjoon Kim, Alla S Kostyukova, Michael R Kreutz, et al. Autism-associated shank3 missense point mutations impact conformational fluctuations and protein turnover at synapses. *Elife*, 10:e66165, 2021.

[129] Marcell Miski, Bence Márk Keömley-Horváth, Dorina Rákóczi Megyeriné, Attila Csikász-Nagy, and Zoltán Gáspári. Diversity of synaptic protein complexes as a function of the abundance of their constituent proteins: A modeling approach. *PLoS computational biology*, 18(1):e1009758, 2022.

[130] Seongsoo Lee, Hsin-Ping Liu, Wei-Yong Lin, Huifu Guo, and Bingwei Lu. Lrrk2 kinase regulates synaptic morphology through distinct substrates at the presynaptic and postsynaptic compartments of the drosophila neuromuscular junction. *Journal of Neuroscience*, 30(50):16959–16969, 2010.

[131] Marco A Pierotti and Angela Greco. Oncogenic rearrangements of the ntrk1/ngf receptor. *Cancer letters*, 232(1):90–98, 2006.

[132] Lorenza Dall'Aglio, Cathryn M Lewis, and Oliver Pain. Delineating the genetic component of gene expression in major depression. *Biological psychiatry*, 89(6):627–636, 2021.

[133] Hyoung-gon Lee, Gemma Casadesus, Akihiko Nunomura, Xiongwei Zhu, Rudy J Castellani, Sandy L Richardson, George Perry, Dean W Felsher, Robert B Petersen, and Mark A Smith. The neuronal expression of myc causes a neurodegenerative phenotype in a novel transgenic mouse. *The American journal of pathology*, 174(3):891–897, 2009.

[134] Zekun Yin, Haidong Lan, Guangming Tan, Mian Lu, Athanasios V Vasilakos, and Weiguo Liu. Computing platforms for big biological data analytics: perspectives and challenges. *Computational and structural biotechnology journal*, 15:403–411, 2017.

[135] Teresa K Attwood, Bora Agit, and Lynda BM Ellis. Longevity of biological databases. *EMBnet. journal*, 21:803, 2015.

[136] Jay E Brenman, Daniel S Chao, Stephen H Gee, Aaron W McGee, Sarah E Craven, Daniel R Santillano, Ziqiang Wu, Fred Huang, Houhui Xia, Matthew F Peters, et al. Interaction of nitric oxide synthase with the postsynaptic density protein psd-95 and α1-syntrophin mediated by pdz domains. *Cell*, 84(5):757–767, 1996.

[137] Marika V Doucet, Andrew Harkin, and Kumlesh K Dev. The psd-95/nnos complex: new drugs for depression? *Pharmacology & therapeutics*, 133(2):218–229, 2012.

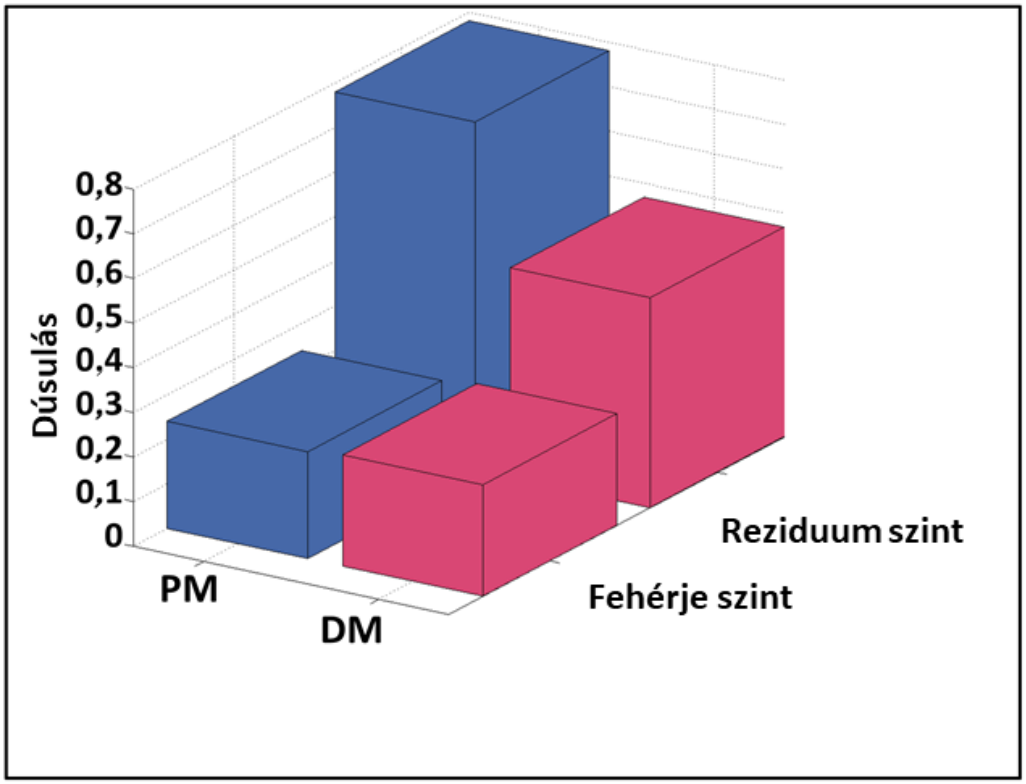
DOI:10.15774/PPKE.ITK.2023.001

[138] Christoph Kleinschnitz, Stine Mencl, Pamela WM Kleikers, Michael K Schuhmann, Manuela G López, Ana I Casas, Bilge Sürün, Andreas Reif, and Harald HHW Schmidt. Nos knockout or inhibition but not disrupting psd-95-nos interaction protect against ischemic brain damage. *Journal of Cerebral Blood Flow & Metabolism*, 36(9):1508–1512, 2016.

[139] Seyed Ziaeddin Alborzi, Amina Ahmed Nacer, Hiba Najjar, David W Ritchie, and Marie-Dominique Devignes. Ppidomainminer: Inferring domain-domain interactions from multiple sources of protein-protein interactions. *PLoS Computational Biology*, 17(8):e1008844, 2021.

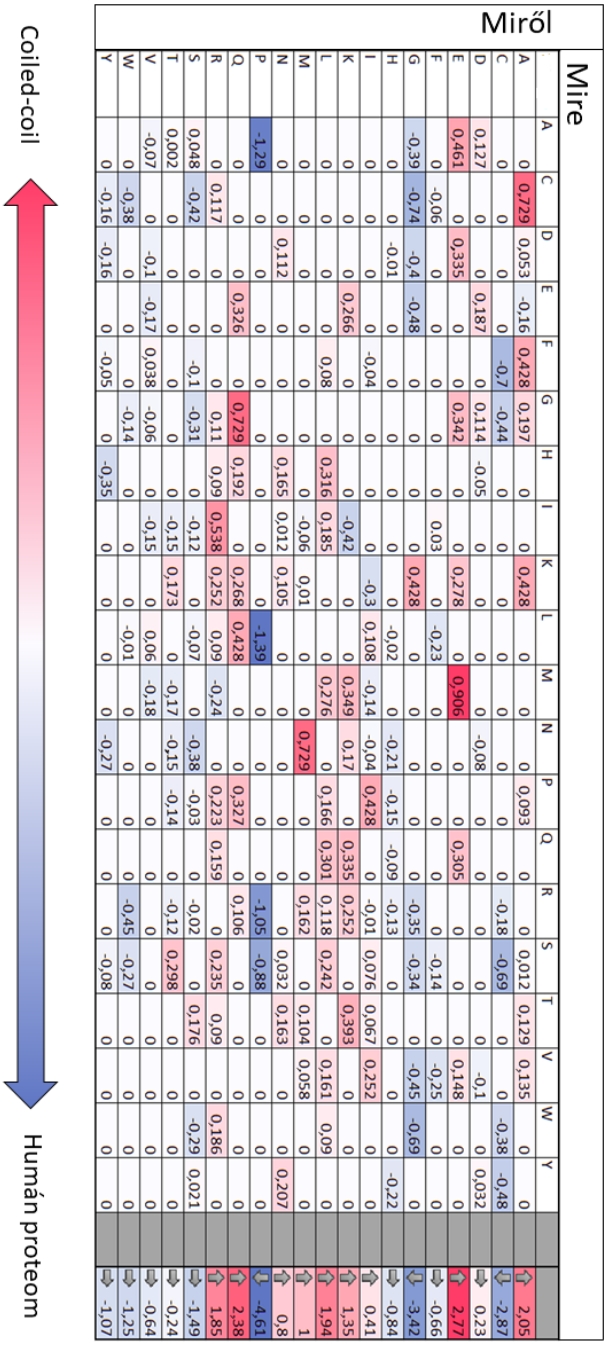
DOI:10.15774/PPKE.ITK.2023.001

10 Függelék



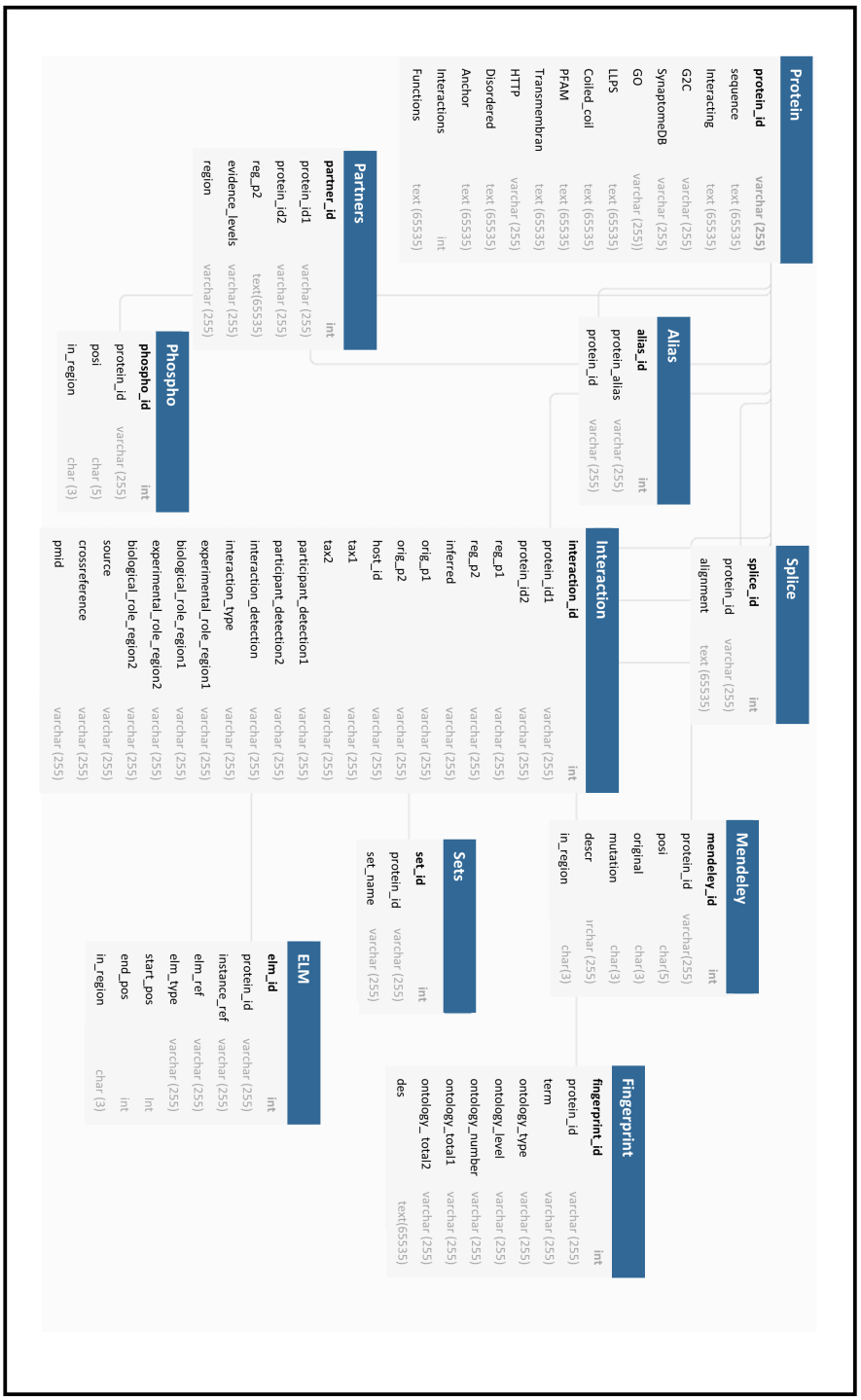
Függelék 1. ábra A DM-ek és a coiled-coilok kapcsolata. A dúsulást (a coiled-coilok és nem coiled-coilok relatív gyakoriságának arányát) reziduum szinten (coiled-coil reziduumok) és fehérje szinten (coiled-coilt tartalmazó fehérje) számítottuk. PM-ek: kék; DM-ek: piros

DOI:10.15774/PPKE.ITK.2023.001



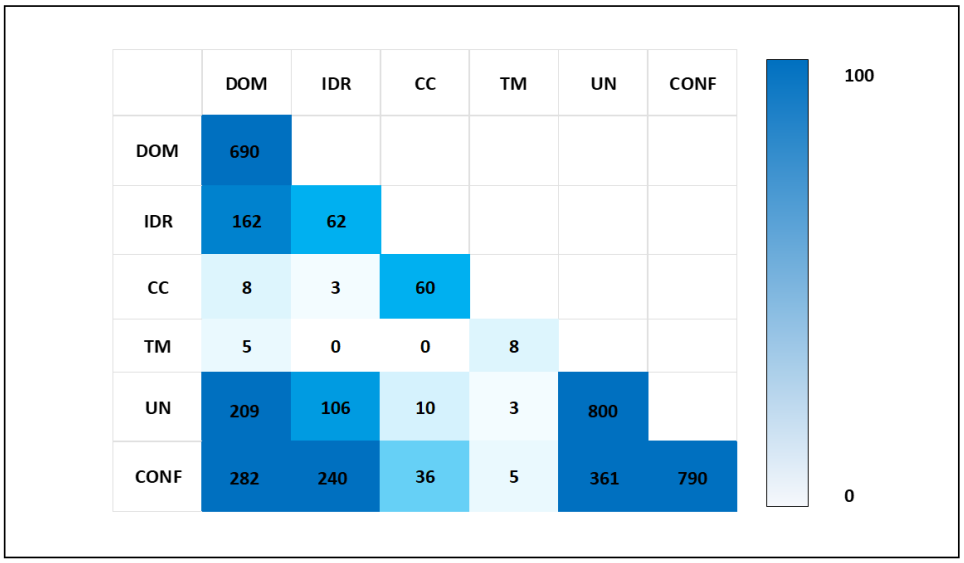
Függelék 2. ábra Egyedi aminosavak kicserélődése a DM-ek által. A coiled-coilra jellemzőbbek pirossal, a humán proteomra jellemzőbbek kékkel

DOI:10.15774/PPKE.ITK.2023.001



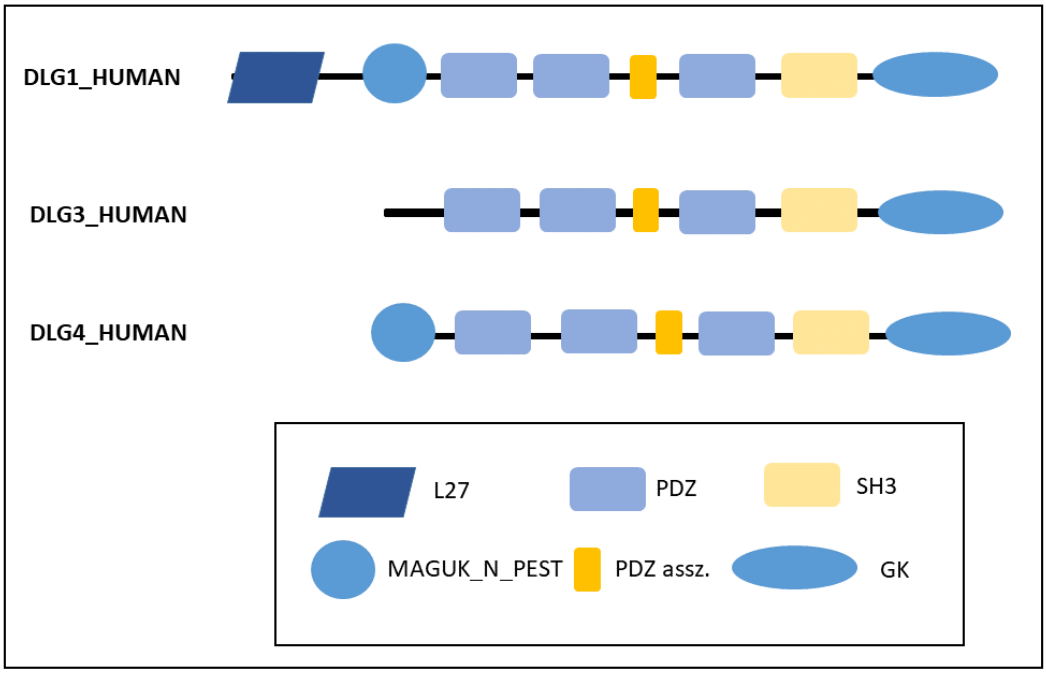
Függelék 3. ábra A PSINDB SQL adatbázis sémája

DOI:10.15774/PPKE.ITK.2023.001

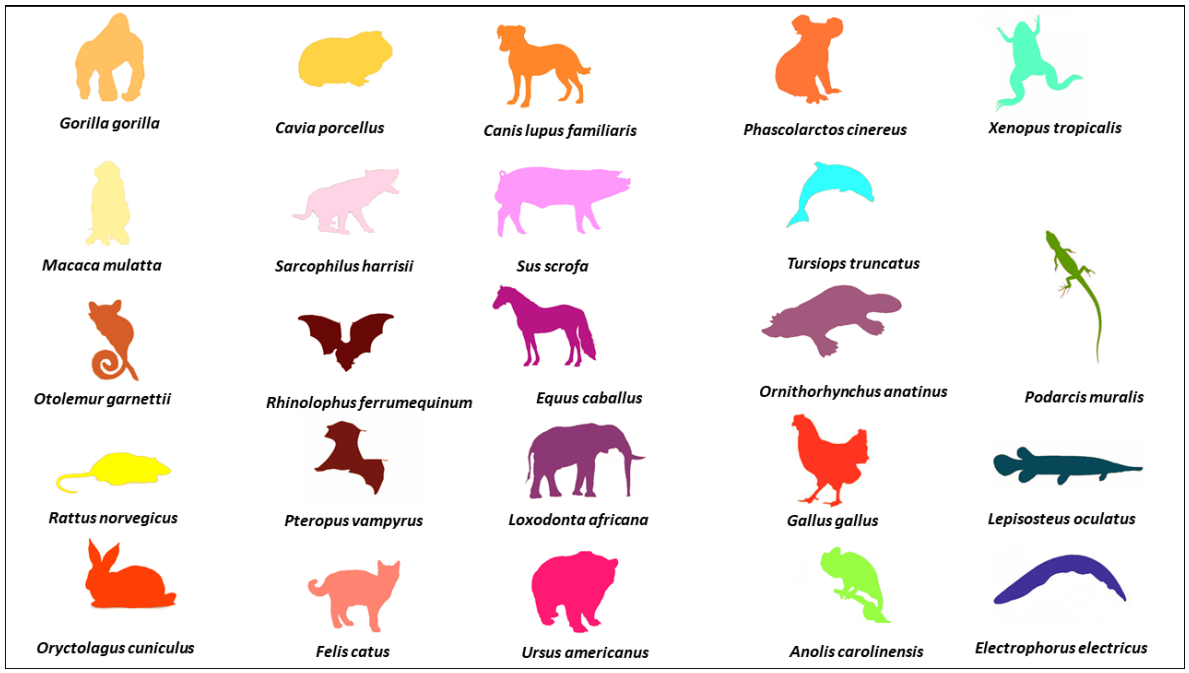


Függelék 4. ábra A kötőrégiók szerkezeti eloszlása a 'suffcient binding regions' szintjén. A jellemzőbb szerkezeti kapcsolatok - mint a domén-domén és domén-rendezetlen - sötétebb színnel találhatóak meg (dom: domén, idr: rendezetlen, cc: coiled-coil,TM: transmemrbán, un: ismeretlen conf: ellentmondádsos)

DOI:10.15774/PPKE.ITK.2023.001

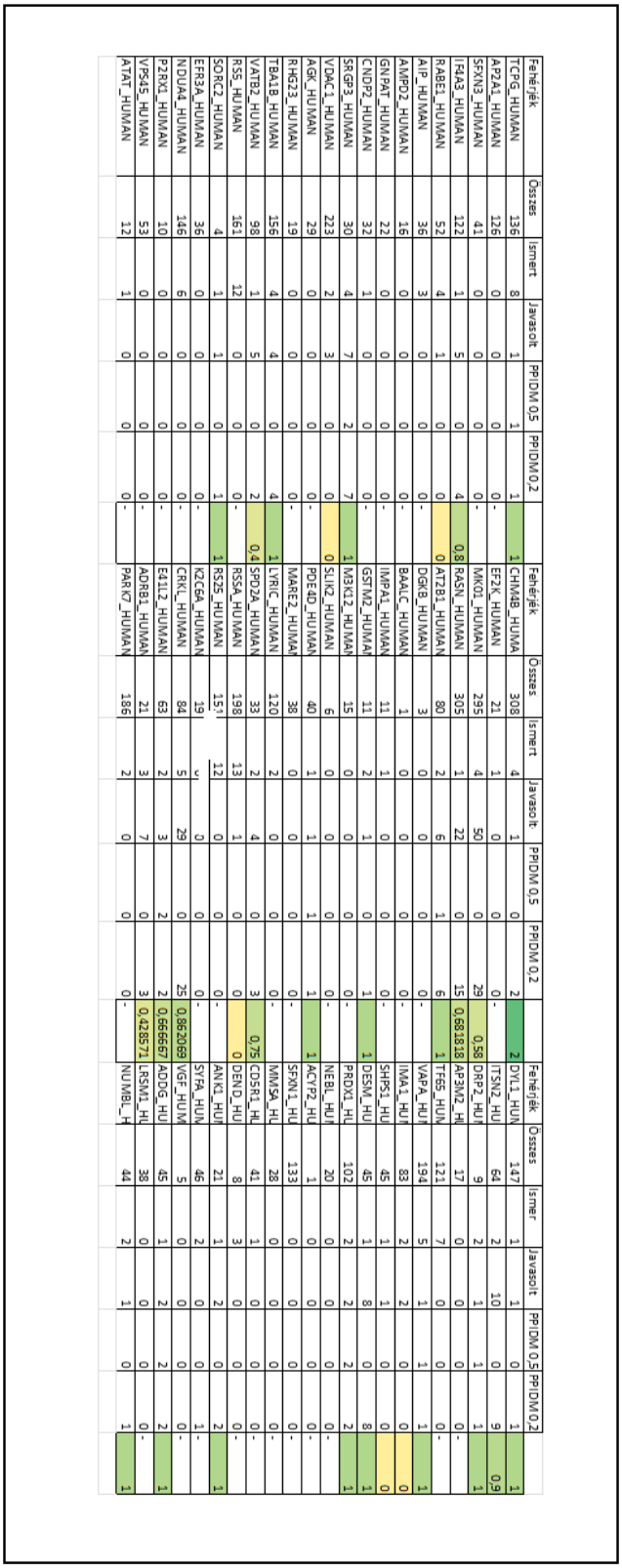


Függelék 5. ábra A humán Dlg fehérjék domén összetétele - számos interakció kialakítására képes egységgel



Függelék 6. ábra A motívum vizsgálatok többszörös szekvencia illesztéshez figyelembe vett fajok (OMA)

DOI:10.15774/PPKE.ITK.2023.001



Fehérjék	Összes	Ismert	Javasolt	PPIDM 0.5	PPIDM 0.2	Fehérjék	Összes	Ismert	Javasolt	PPIDM 0.5	PPIDM 0.2	Fehérjék	Összes	Ismer	Javasolt	PPIDM 0.5	PPIDM 0.2
TCPG_HUMAN	136	8	1	1	1	CHM4B_HUMA	308	4	1	0	2	DYL1_HUN	147	1	1	1	1
AP2A1_HUMAN	126	0	0	0	0	EF2K_HUMAN	21	1	0	0	0	ITSN2_HU	64	2	10	0	0
SFXN3_HUMAN	41	0	0	0	0	MK01_HUMAN	295	4	50	0	29	DRP2_HUM	9	2	1	1	0
IFA3_HUMAN	122	1	5	0	4	RASN_HUMAN	305	1	22	0	15	AP3M2_HU	17	0	0	0	0
RABE1_HUMAN	52	4	1	0	0	AT2B1_HUMAN	80	2	6	1	6	TFG5_HUMA	121	7	0	0	0
AIP_HUMAN	36	3	0	0	0	DGKB_HUMAN	3	0	0	0	0	VAPA_HUM	194	5	1	1	1
AMPD2_HUMAN	16	0	0	0	0	BAALC_HUMAN	1	0	0	0	0	IMA1_HUM	83	2	2	0	0
GNPAT_HUMAN	22	0	0	0	0	IMPA1_HUMAN	11	1	0	0	0	SHPS1_HU	45	1	1	0	0
CNDP2_HUMAN	32	1	0	0	0	GSTM2_HUMA	11	2	1	0	1	DESM_HU	45	1	8	0	1
SRGP3_HUMAN	30	4	7	2	7	M3K12_HUMAN	15	0	0	0	0	PRDX1_HU	102	2	2	2	2
VDAC1_HUMAN	223	2	3	0	0	SLIK2_HUMAN	6	0	0	0	0	NEBL_HUM	20	0	0	0	0
AGK_HUMAN	29	0	0	0	0	PDE4D_HUMAN	40	1	1	1	1	ACYP2_HU	1	0	0	0	0
RHG23_HUMAN	19	0	0	0	0	MARE2_HUMAN	38	0	0	0	0	SFXN1_HU	133	0	0	0	0
TBA1B_HUMAN	156	4	4	0	1	LYRIC_HUMAN	120	2	0	0	0	MM3A_HU	28	0	0	0	0
VATB2_HUMAN	98	1	5	0	0,4	SPD2A_HUMAN	33	2	4	0	3	CD5R1_HU	41	1	0	0	0
RS5_HUMAN	161	12	0	0	0	RSSA_HUMAN	198	13	1	0	0	DEND_HU	8	3	0	0	0
SORC2_HUMAN	4	1	1	0	1	RS2S_HUMAN	15	12	0	0	0	ANKS_HUM	21	1	2	0	2
EFR3A_HUMAN	36	0	0	0	0	K2C6A_HUMAN	19	5	0	0	0	SYFA_HUN	46	2	0	0	0
NDUA4_HUMAN	146	6	0	0	0	CRKL_HUMAN	84	5	29	0	25	VGF_HUM	5	0	0	0	0
P2RX1_HUMAN	10	0	0	0	0	E41L2_HUMAN	63	2	3	2	2	ADDG_HU	45	1	2	2	1
VPS45_HUMAN	53	0	0	0	0	ADRB1_HUMAN	21	3	7	0	3	LRSM1_HU	38	0	0	0	0
ATAT_HUMAN	12	1	0	0	0	PARK7_HUMAN	186	2	0	0	0	NUMBL_H	44	2	1	1	1

Függelék 7. ábra Részlet a domén-domén kötőrégiók vizsgálata során kapott eredmények és egy korábban felállított és publikált módszer, a PPIDomainMiner [139] eredményeinek összehasonlításából (A PPIDM esetében két tresholddal)