

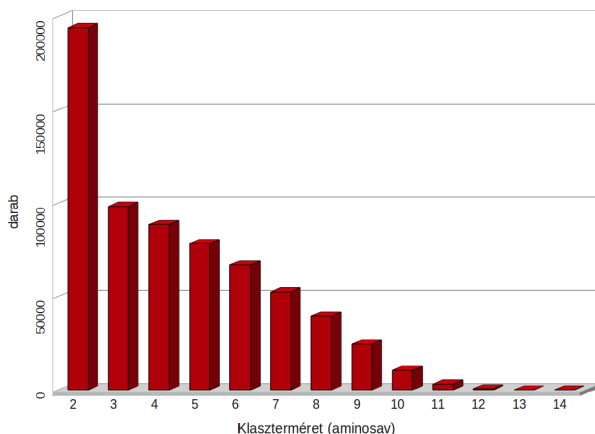
## Szakmai zárójelentés

a 75460 sz. „Transzmembán fehérjék elméleti szerkezet vizsgálata és tudásbázisa” című OTKA pályázatról

A pályázat során – a munkatervnek megfelelően – elkészítettük a TOPDB adatbázis új web interfészét meghajtó programkönyvtárat (XBuilder). Ez, egy C++ nyelven írt program könyvtár, amelyhez a szintén C++ nyelven írt web toolkit (wt, <http://www.webtoolkit.eu/wt>) könyvtárat használtuk. Ez az interfész a modern programozói követelményeknek megfelel, ugyanis az interfészben teljesen szeparáltuk a modellt, a megjelenés és a felhasználók tevékenységét megadó egységeket (MVC, model-view-controller). Az XBuilder programcsomag lehetővé teszi, hogy egy XML paraméter fájlban megadott adatok alapján teljesen automatikusan épüljön fel és jelenjen meg egy dinamikus web oldal. Természetesen az elkészült Xbuilder programkönyvtárat önmagában nem közölhettük, mint tudományos eredményt, a könyvtárat felhasználó publikus és tudományos eredményeket, szervereket és adatbázisokat leíró eredményeket igen. Bár a TOPDB adatbázis új webfelületén még dolgozunk, és jelenleg még nem publikáltuk, az XBuilder könyvtárat használtuk egy új projekt megvalósítása során is (lásd lalább), illetve az ebben a programkönyvtárban levő lehetőségeket kívánjuk használni a PDBTM új webfelületének a létrehozásánál is.

A pályázat első évben módosítottuk a HMMTOP transzmembrán topológia becslő algoritmust oly módon, hogy a meghatározott topológiai kísérleti adatokat, még ha azok ellentmondóak is, figyelembe tudja venni a topológia becslés során. Ehhez mind a optimalizáció során használt Baum-Welch algoritmust, mind a becsléshez használt Viterbi algoritmus kódját módosítani kellett. A HMMTOP algoritmusról, annak helyes használatáról, valamint a módosításokról az ebben az időben készült, felkérésre írt könyvfejezetben részletesen is írtunk (Tusnady GE and Simon I: ***Shedding light on transmembrane topology***, Protein Structure Prediction: Method and Algorithms , 107-136., 2010). A módosított algoritmust – egy amerikai csoporttal együttműködésben - sikeresen alkalmaztuk egy eddig még ismeretlen topológiájú fehérje, a SLC20A1 jelű foszfát transzporter fehérje topológia becsléséhez. Ez a III-as típusú nátrium-függő inorganikus foszfát transzporter azért került a kutatók nézőterébe, mert kiderült, hogy számos gamma-retrovírus ezt a transzportert használja a sejthez való kötődéshez, ami az első lépése a fertőzésnek. Emiatt, nemcsak molekuláris szempontokból érdekes ezen fehérje vizsgálata, hanem virológiai szempontokból is. Ebben a munkában először kapcsoltuk össze a kísérletek tervezését, kivitelezését az elméleti topológia becslés lehetőségeivel. Az erről készült munkánk a J. Biol. Chem. folyóiratban publikáltuk (Farrell KB, Tusnady GE and Eiden MV: ***A new structural arrangement of the extracellular regions of the phosphate transporter SLC20A1, the receptor for gibbon ape Leukemia virus.***, J Biol Chem 284, 29979-87., 2009).

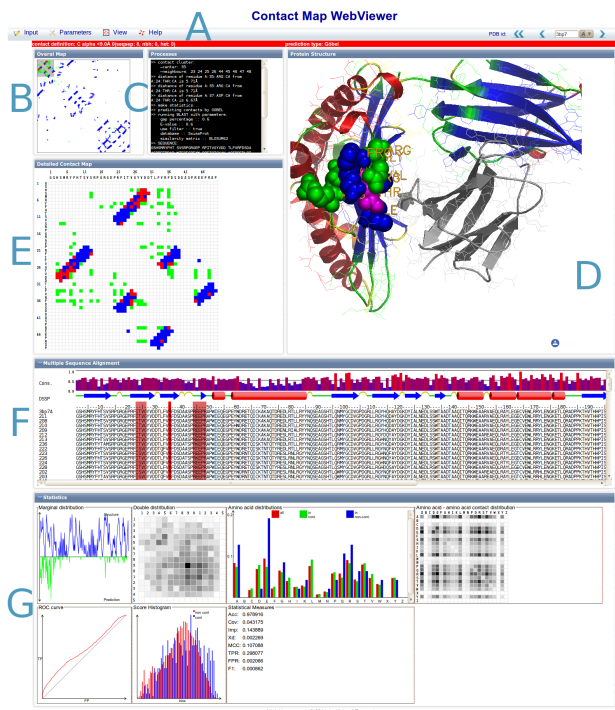
A transzmembrán fehérjék membránban való pontos helyzetének a meghatározásához szükséges új eljárás kidolgozása közben számos problémába ütköztünk. Ezek közül az egyik az volt, hogy kiderült, a



modellizéshez használni kívánt APBS (Adaptive Poisson-Boltzmann Solver) módszer parametrizációja membrán fehérjékre nem használható. Új paraméterek tesztelésére nekünk nem volt lehetőségünk, ráadásul közben megjelent egy cikk (Callenberg et al, APBSmem: a graphical interface for electrostatic calculations at the membrane. PLoS One. 2010 Sep 29;5(9). pii: e12722.), amelyben ezt a problémát nagyon részletesen leírták, és megoldották. A cikkben leírt apbsmem szoftver TMDet integrációja jelenleg még folyamatban van.

1. ábra: Kontaktusban levő aminosavak klaszterméret eloszlása globuláris fehérjékben.

A transzmembrán fehérjék szerkezet becsléséhez, illetve a transzmembrán szakaszok közötti kölcsönhatások becsléséhez az ún korrelált mutációs eljárás alapján kontaktus becslő módszert kívántunk felhasználni. Ezért először összegyűjtöttünk több kontaktus becslő eljárást, és teszteltük ezek becslési pontosságát mind globuláris, mint transzmembrán fehérjéken. Azt tapasztaltuk, hogy bár a valódi kontaktusok feldúsulnak a becslt kontaktusok között, a becslés pontossága, még ha az ún  $\delta-4$  módszert alkalmazzuk a kiértékelés során, sem éri el azt a szintet, ami lehetővé tenné ezen módszerek transzmembrán fehérjék membránba ágyazott részeinek pontos szerkezeti becslését. Megvizsgáltuk, hogy mi lehet a szerkezeti/szekvenciális oka, ezen módszerek igen alacsony becslési pontosságának. Az irodalomban számos okot találhatunk, pl nem megfelelő szekvencia illesztés használata, nem megfelelő scoring mátrix használata, „evolúciós zaj” stb., azonban általunk generált mesterséges szekvenciák és szekvencia illesztéseken végzett becslések segítségével ezeket mind ki tudtuk zárni. Mivel a fehérjékben levő aminosavak koordinációs száma kettőnél nagyobb (1. ábra), ezért nem várható, hogy egy konkrét aminosav változását a térben mindig csak egy szomszédja kompenzálja, hanem a térbeli szomszédok közül bármelyik kompenzálhatja. Véleményünk szerint a kontaktus becslés alacsony teljesítőképességének ez az alapvető oka, mert ezekben az aminosav párok korrelációját vizsgálják a szekvencia illesztésben, holott nagyobb klaszterek együttes változását kellene figyelembe venni. Jelenleg egy ilyen módszeren dolgozunk, amely képes magasabb rendű korrelációkat figyelembe venni a szerkezet becslés során.



2. ábra: CMWeb szerver. A, Menü és navigáció; B, Teljes kontaktus térkép; C, Információs mező; D, Szerkezet nézegető; E, Nagyított kontaktus térkép részlet; F, Szekvencia illesztés; G, statisztikai adatok.

A kontaktusok vizsgálatát nagyban megkönnyíti, ha a kontaktus térképen levő kapcsolatokat mind a fehérjék térszerkezetében, mind a szekvencia illesztésben egyszerre tudjuk vizsgálni, vizualizálni. A projekt során kidolgozott Xbuilder könyvtár segítségével ezért létrehoztunk egy web szervert, amely az interneten keresztül interaktív módon képes egy-egy fehérje 3D-s szerkezetét megjeleníteni, a kontaktus térképpel, illetve szekvencia illesztéssel együtt. A Szerver nemcsak megjeleníti ezeket az információkat, hanem azokat interaktív módon össze is kapcsolja. Így a kontaktus térképen egy adott kontaktus könnyen megjeleníthető a 3D-s szerkezetben, illetve a szekvencia illesztésben (2. ábra). A szervert CMWeb szervernek neveztük el, és publikussá tettük az akadémia felhasználók számára (<http://cmweb.enzim.hu>). A CMWeb szerverről szóló közleményünk jelenleg közlés alatt van a Nucleic Acids Research folyóiratban (Kozma D, Simon I and Tusnády GE (2012), CMWeb: an interactive on-line tool for analysing residue - residue contacts and contact prediction methods. Nucl Acids Res, Web Server issue).

A projekt 3 esztendeje alatt több könyvfejezet, illetve összefoglaló cikk megírására kaptunk felkérést a transzmembrán fehérjék topológia becslésével illetve szerkezeti adatainak összegyűjtésével kapcsolatban (Tusnády GE and Simon I (2010), Shedding light on transmembrane topology. Protein Structure Prediction: Method and Algorithms , 107-136; Tusnády GE and Simon I (2010) Resource for

structure related information on transmembrane proteins. Structural Bioinformatics of Membrane Proteins , 45-59; Tusnady GE and Simon I (2010) Topology prediction of helical transmembrane proteins: how far have we reached? Curr Protein Pept Sc 11, 550-61).

Az elmult evben, egy tovabbi az OTKA kutatas tamogatasaval megjelent egy cikkunk a Epigenetics címu folyoiratban, amelyben a belgiumi Center for Medical Genetics (Ghent University Hospital) munkatarsaival egyutt, egy új, biszulfittal kezelt DNS mintakon használható PCR primer tervezési algoritmusrol szolo implementaciot irtunk le (Lefever S, Hoebeeck J, Pattyn F, Tusnady GE, De Paepe A, Speleman F, Aranyi T and Vandesompele J (2010) methGraph: A genome visualization tool for PCR-based methylation assays. Epigenetics 5, 159-63).