

A SOKSZÍNŰ FEHÉRJE¹

THE MULTIFORM PROTEIN

Simon István

az MTA levelező tagja

MTA Természettudományi Kar Enzimológiai Intézet

simon.istvan@ttk.mta.hu

Kulcsszavak: bioinformatika, számítógépes fehérjeszerkezet-bebecslés, vizes közegben oldódó rendezett fehérjék, transzmembrán fehérjék, eredendően rendezetlen fehérjék, fehérje-fehérje komplexek

Keywords: protein structure, bioinformatics approaches, structure predictions, web servers, transmembrane protein, disordered protein

Akadémiai bemutatkozó előadásomban magamból a kutatót kell bemutatni. A kutatót eredményei azonosítják, tehát ezekből szedtem össze egy csokorra valót.

Pályám kezdetén, 1969-ben Magyarországon a fehérjeszerkezet-kutatás egyértelműen kísérletes munkát jelentett. Én is ezt végeztem *Elődi Pál* csoportjában. Többek között bevezettem Magyarországon a fehérjék kisszögű röntgen-szórásos vizsgálatát és kifejlesztettem egy hatékony módszert kis szerkezetváltozások detektálására (Simon, 1971). Ezzel azután a következő két évtizedben a fehérjékkel kapcsolatban elég sok kérdést sikerült megválaszolni. A kandidátusi fokozat 1975-ös megszerzése után a Cornell Egyetemen *Harold A. Scheraga* csoportjában töltött fél éves posztdoktori tanulmányutam során kidolgoztam egy, a polimerek szerkezetének számítógépes meghatározására szolgáló, később *build up* néven ismertté vált módszert (Simon et al., 1978). Hazatérve, az akkor már *Závodszy Péter* vezette csoportban a kísérletes munkák mellett a fehérjeszerkezet szerveződésének elméleti vizsgálatával is elkezdtem foglalkozni. 1982-ben a minnesotai egyetemen *Clare Woodward* csoportjában egymással kölcsönható fehérjék egymás dinamikus sajátosságaira való hatását sikerült meghatároznom hidrogén–deutérium kicserélődési kinetikák NMR-es (mágneses magrezonanciás) mérésével (Simon et al., 1984). Írtunk egy azóta is rendszeresen hivatkozott

¹ A 2016. december 13-án elhangzott székfoglaló előadás szerkesztett változata, amely a vetített ábrák helyett az előadásban említett munkáimból készült publikációk listáját közli.

áttekintő cikket is (Woodward et al., 1982). Az 1986-os tanulmányutam Scheraga csoportjába vezetett vissza, ahol a már említett build up módszerrel meghatároztam a természetes cellulóz atomi felbontású térszerkezetét, amit oldhatatlansága miatt akkor még kísérletesen nem lehetett meghatározni (Simon et al., 1988a, 1988b). A tudomány doktora fokozat megszerzése és saját kutatócsoportom 1987-es alapítása után, addigi elméleti munkásságom megkoronázása volt, hogy 1989–1990-ben újra Scheraga professzor csoportjába kerülve megmutattam, hogy egy fehérje térszerkezete pusztán számítógépes eljárással meghatározható a kémiai szerkezet ismeretében (Simon et al., 1991). Természetesen itt is a build up módszert használtam, azaz a polipeptidlánc rövid átfedő szegmenseinek valamennyi alacsony energiájú konformációját kombinálva választottam ki a teljes szerkezet felépítéséhez használható elemeket. Ez a munka mérföldkő, de inkább csak kultúrtörténeti okból: ez az első atomi szintű, energiaminimalizációs fehérjeszerkezet-számítás.

A szegmensekből való építkezés csak a legfontosabb kölcsönhatásokat, a szekvenciálisan közeli aminosavak között lévőket veszi figyelembe. Módszereünk javítása céljából a nem kovalens keresztkötéseket biztosító kölcsönhatási csomópontokra diákjaimmal, *Dosztányi Zsuzsannával* és *Fiser Andrással* bevezettük a stabilizációs centrumok fogalmát, amelyet széles körben használnak a fehérjetudományban. Kidolgoztunk egy módszert is a centrumok aminosavainak a szekvenciából való becslésére (Dosztányi et al., 1997). Definíciónk csak a kölcsönhatások számát tekintette, energiájukat nem. Ugyanakkor meg voltunk róla győződve, hogy ezeknek a centrumoknak közül van a fehérjék stabilitásához. Végül, húsz évvel a stabilizációs centrumok bevezetése után, sikerült bizonyítani munkatársammal, *Magyar Csabával*, hogy a stabilitási centrumok száma szoros korrelációban van a fehérjék hőstabilitásával (Magyar et al., 2016). Ennyit pályafutásom első, hosszabbik részéről.

Az igazán érdekes időszak azonban csak ezután kezdődött. Két évtizede óriási szerencsémre egymástól függetlenül két nagy tudományos áttörés történt, amelyek közül az egyik az igényt, a másik a lehetőséget teremtette meg a *bioinformatika* létrejöttéhez. A genom projektek az élettudományok területén eddig nem látott adatmennyiségeket eredményeztek, amelyek feldolgozása lehetetlen lett volna megfelelő számítástechnikai háttér nélkül. Ezzel egyidejűleg jelent meg az *internet*, lehetővé téve az adatforrások összekapcsolását és az adatok feldolgozását. Látszólag csak technikai, de valójában alapvető változás volt, hogy a fehérjék aminosavsorrendjének meghatározásában a fehérjeszintű vizsgálatokról áttértek a fehérjéket kódoló nukleinsavak bázissorrendjének meghatározására és ebből azonosították az aminosavsorrendeket. Így azoknak a fehérjéknek is megismerték az aminosavsorrendjét, amelyeket nehéz lett volna tiszta állapotban előállítani és vizsgálni. Tipikusan ilyenek a membránba ágyazott fehérjék, illetve a rendezetlen fehérjék, amelyek létezéséről ekkor még nem is tudtunk.

Az elméleti munkák természetesen a kísérletesen nehezen vizsgálható fehérjékre fókuszáltak. Ilyenek a membránokon átnyúló, ún. transzmembrán fehérjék. Ezek a fontos fehérjék biztosítják az anyag- és információáramlást mind a sejtek alkotóelemei között, mind a sejt és környezete között. Ahhoz, hogy tiszta állapotban előállítsák, különösen, hogy röntgendiffrakciós célra egykristályt készítsenek belőle, ki kell vonni a membránból. Ilyenkor az eddig a membránhoz illeszkedő hidrofób felszíni részek egymáshoz tapadnak és a fehérje kicsapódik az oldatból. Ennek megfelelően a jelenleg ismert több mint százezer fehérje-térszerkezet közül az összes fehérje 25–30 százalékát kitevő transzmembrán fehérjéből csak kb. kétezer szerkezet ismert. Ráadásul a preparálásakor elvesz egy igen fontos információ, hogy hogyan helyezkedett el eredetileg a fehérje a membránban. A membrán belsejében nincs víz vagy más protondonor, illetve akceptor, ami felszakíthatná a peptidváz hidrogénhidjait, ezért a membrán belsejében általában eltolási szimmetriájú szerkezeteket találunk. Ezek gyakorlatilag alfa-hélixek és csak öt százaléknyi speciális esetben béta-szerkezetek. *Tusnády Gáborral* és *Dosztányi Zsuzsával* kidolgoztunk egy algoritmust, a TMDet nevűt, amellyel a fehérje atomi koordinátáiból, az aminosavak hidrofób jellegéből és egyéb fizikai-kémiai adatokból számítással meghatározható a membrán legvalószínűbb helyzete és orientációja (Tusnády et al., 2004, 2005a). Ennek felhasználásával létrehoztuk a PBDTM transzmembrán térszerkezeti adatbázist, amely a közhasználatú fehérjeadatbank, a PDB adataira építve már tartalmazza a membránra vonatkozó információkat is (Tusnády et al., 2005b). Ezt a közelmúltban jelentősen korszerűsítettük (Kozma et al., 2013).

Sok gyakorlati feladat megoldásához az atomi felbontású térszerkezet helyett elegendő a polipeptidlánc topológiájának ismerete, azaz annak ismerete, hogyan megy oda-vissza a lánc a membrán két oldala között és melyek a membránban található láncresek. Kísérletes módszerekkel csak többhetes vagy hónapos munkával lehet a topológiát megadni. Ezért olyan nagy az igény a szekvenciából elméleti módszerekkel történő topológiabecslésre. Mi az elsők között, még a kilencvenes években kapcsolódtunk be ebbe a munkába, lényegében véletlenül. Nem egy becslő módszert akartunk csinálni, hanem amikor ismertté vált az egyik első transzmembrán fehérje térszerkezete, nevezetesen a bakteriorodopsziné, doktoranduszommal, *Cserző Miklóssal* és külföldi kollégákkal megpróbáltuk homológiamodellezéssel meghatározni az emberi szem rodopszinjának szerkezetét. Ez egy rutin eljárás, ha az ismert térszerkezetű fehérjék között találunk olyat, amelynek a szekvenciája sok helyen hasonlít a mi ismeretlen térszerkezetű fehérjénkére. Azon az alapon, hogy a hasonló szekvenciák hasonló szerkezetekbe rendeződnek, meg lehet találni a keresett szerkezetet.

Ez azonban nem sikerült, mert a két fehérje között kevesebb mint 20% a szekvenciális azonosság. Azért nem tudtuk a szekvenciákat egymáshoz illeszteni, mert a két fehérjének a nevükön kívül szinte semmi közük nincs egymáshoz.

Úgy jártunk, mintha nefelejcsvirágból memóriajavító gyógyszert próbáltunk volna készíteni. Viszont kiderült, hogy egy, korábban diákjaimmal, *Tüdős Évával* és Cserző Miklóssal fehérjetervezés céljára készített aminosav-hasonlósági mátrixunk alkalmazásával a hagyományos aminosavsorrend-illesztési eljárásban az összes transzmembrán szakasz rokonként ismeri fel egymást, függetlenül attól, hogy a transzmembrán fehérjék milyen szekvenciáit hasonlítjuk össze (Tüdős et al., 1990). Ez lehetőséget biztosított a topológia becslésére. A DAS-algoritmusból egy 43 transzmembrán fehérje szekvenciájából álló könyvtárral vetjük össze az ismeretlen topológiájú transzmembrán fehérje szekvenciáját, és a természetesen nem valódi homológiaképek átlaga jelöli ki a transzmembrán szakaszokat (Cserző et al., 1997). A módszert később továbbfejlesztettük, hogy a szekvencia alapján felismerje, valódi transzmembrán fehérjét vizsgálunk-e (Cserző et al., 2002, 2004). Erre a korabeli algoritmusok még nem voltak alkalmasak.

Később megfejtettük a transzmembrán fehérjék topológiaszerveződésének elméleti hátterét is. A különböző közegben eltérően oldódó anyagokra érvényes törvényszerűségekből kiindulva egy transzmembrán fehérje esetén a polipeptidlánc legvalószínűbb topológiáját az tünteti ki, hogy abban a különböző fizikai-kémiai tulajdonságú térrészekben található láncrészek összesített aminosav-kompozíciója maximálisan eltér egymástól. Kiváló doktoranduszom, Tusnády Gábor, algoritmusában ilyen térrésznek tekintette a membrán belsejét, a sejt belseje és külseje felé eső határfelületeket, a sejt belsejében, illetve a sejten kívül található szabad folyadékot. A térrészekben található kompozíciók maximális eltérése mint szélsőérték-probléma egy ún. rejtett Markov-modell segítségével numerikusan megoldható (Tusnády–Simon, 1998). Ráadásul, az adott eljárás különböző peremfeltételek megadásával is működik. Tehát, ha van valamilyen előzetes ismeretünk a topológiáról, az ehhez a feltételhez tartozó szélsőértéket is meg tudja találni (Tusnády–Simon, 2001). Túl azon, hogy ez a HMMTOP-nak nevezett predikciós módszer egyike a legjobbaknak, legalább ilyen fontos, hogy fizikai képet ad a topológia szerveződéséről. Rámutat, hogy egy entrópiaszerű függvény maximuma tünteti ki a legvalószínűbb topológiát. Számomra, mint alapkutatót végző kutató számára, igazi tudományos értéke a mértékre adott válaszoknak van. Ennyit a transzmembrán fehérjékről.

A genom projektek legmeglepőbb eredménye az ún. eredendően rendezetlen fehérjék létezésének felismerése volt. Kiderült, hogy vannak fehérjék, amelyek egészben vagy egyes láncszakaszaikban nem alakítanak ki stabil térszerkezetet, hanem nagyszámú szerkezet dinamikusan egyensúlyában léteznek. Az ilyen fehérjék a szokásos fehérjevizsgálati módszerek számára láthatatlanok. Forralás hatására nem csapódnak ki, denaturáló szerek hatására nem mutatnak elváltozást, gyakorlatilag nem nyelik el a 220 nm hullámhosszú ultraibolya sugarakat. Később kiderült, hogy nemcsak jelentős mennyiségben léteznek, de még röntgen-diffrakciós kép is készíthető róluk, igaz, hogy csak makromolekulákkal, például

fehérjékkal alkotott komplexeikről. Ezekben a komplexekben a rendezetlen fehérje szerkezete attól függ, hogy mihez kötődött. Tehát nem igaz az a paradigma, hogy egy fehérje térszerkezetét mindig annak aminosavsorrendje határozza meg.

Mi is, mint sokan mások, először arra voltunk kíváncsiak, mitől lesz rendezetlen egy fehérje vagy egy fehérjerészlet. A nyilvánvalóan sokféle és láthatatlan rendezett szerkezet helyett célszerűbbnek látszott a rendezett fehérjék közös sajátosságainak azonosítása, és amelyek fehérjének nincs ilyen, azt rendezetlennek kell tekinteni. Az aminosavak közötti kölcsönhatások energiáját a húszféle aminosav összes lehetséges párjára meg lehet becsülni abból, hogy a párok az egyedi aminosavak gyakoriságához képest milyen gyakran kerülnek egymás közelébe az ismert térszerkezetű fehérjékben. Ezekhez a gyakorisági értékekhez a Boltzmann-törvény energiaértékeket rendel, amelyekkel a térszerkezet alapján egy fehérje teljes párkölcsönhatási szabadenergiáját meg lehet határozni. Ezt a szabadenergia-értéket a gyakorlatban arra használják, hogy meghatározzák egy ismeretlen térszerkezetű fehérjéről, hogy hasonlít-e egy már ismert térszerkezetűre. Ugyanis, ha az ismert szerkezetekben az aminosavakat kicseréljük a mi fehérjénk aminosavaival, általában kedvezőtlenül magas energiaértékeket kapunk. Kivéve azt az egy szerkezetet, amelyikbe a mi fehérjénk fel tud úgy tekeredni, hogy alacsony energiájú és stabil szerkezetű legyen. Kiváló munkatársnőm, korábbi diákom, Dosztányi Zsuzsa megvizsgálta, mekkorák ezek a térszerkezetekhez tartozó szabadenergia-értékek és kiderült, hogy nagyszámú fehérjére megnézve ez az érték egy-egy aminosavra számolva igen hasonló (Dosztányi et al., 2005a).

Úgy látszik, ennyire van szükség, hogy egy fehérje stabil maradjon, ugyanakkor meglegyen például az enzimek működéséhez elengedhetetlen flexibilitása. Ez arra utal, hogy egyes fehérjék, illetve fehérjerészletek azért rendezetlenek, mert a kölcsönhatások nem fedezik azt az energiamennyiséget, amennyi a rendeződéskor fellépő entrópiavesztés hőmérséklettel vett szorzatát kompenzálná. Az aminosavpárok kölcsönhatási energiáit optimalizálандó paramétereknek tekintve nagyszámú fehérjére felírható a becsült párkölcsönhatási energia kifejezése. Ezeket összevetve a vizsgált fehérjék tényleges szerkezetéből számolt energiákkal, megkapjuk a kölcsönhatási paraméterek értékeit. Ezekkel az optimalizált paraméterekkel a térszerkezet ismerete nélkül is becsülhető az egyes fehérjéhez rendelhető párkölcsönhatási energia. Látszik, hogy a térszerkezetekből számolt és a szekvenciákból becsült energiaértékek jól korrelálnak. Ezért dolgozhatunk a becsült értékekkel, amihez nincs szükség a térszerkezetek ismeretére, tehát a becsült rendezetlen fehérjékre is megoldható. Azt tapasztaltuk, hogy a rendezett és rendezetlen fehérjékre számított értékek jól elkülönülnek. Sőt, a rendezetlenségi hajlam meghatározható a vizsgált polipeptidlánc minden aminosavára külön-külön is egy elég nagy láncszakaszon belüli szekvenciából. Ezt a 0–1 tartományra normál függvényt adja ki az algoritmusra épült szervertünk, az IUPred (Dosztányi et al., 2005b). A rendezetlen fehérjék általában regulációs és jelölőfunkciókat

látnak el, értelemszerűen más makromolekulákhoz, főleg más fehérjékhez kötődve. A több ponton kötődés rendeződéssel jár, azaz jelentős entrópiavesztés kompenzálja az atomi kölcsönhatásokból származó entalpianyereséget. A több ponton történő kötődés specifikus, de gyenge, könnyen disszociálódó, mert a kötődés, az azzal járó rendeződés miatt jelentős entrópiavesztéssel jár, így az csak minimális szabadenergia-változással jár. Pontosan ez az, amihez például foszforiláláshoz vagy egyéb jelöléshez szükség van.

A rendezetlen fehérjék működése szempontjából elsődleges kérdés, hogy a rendezetlen szakaszok mely részei tudnak más fehérjéhez kötődni, és annak felszínén rendeződni. Munkatársam, *Fuxreiter Mónika* vette az ELM-adatbázist, amelyik ezeket a kötésben található motívumokat gyűjtötte össze. A motívumok közepénél fogva az összes teljes szekvenciát egymásra helyezte úgy, hogy a motívumközepek mind egy koordináta-rendszer origójára essenek. Majd megnézte az átlagos rendezetlenséget az origótól való távolság függvényében. A kötőhelyek a rendezett és rendezetlen részek határára esnek, tehát várható, hogy minimális extra stabilizáló kölcsönhatásra stabilizálódni tudnak és megvalósul a csatolt kötődés és rendeződés jelensége (Fuxreiter et al., 2007). Dosztányi Zsuzsa és doktoranduszunk, *Mészáros Bálint* megnézték, hogy melyek azok a polipeptidlánc-szakaszok, amelyek aminosavai egymással nem tudnak a rendeződéshez elegendő energiájú kölcsönhatást kialakítani, de egy rendezett globuláris fehérje aminosavaival már igen. Ehhez vettek sok fehérjét a PDB-ből, és ezekből meghatározták az átlagos aminosav-kompozíciót. Ezek után rendezetlen kötőhelyet ott várunk, ahol az aktuális szekvenciális környezet aminosav-összetételét erre az átlagos globuláris fehérje aminosav-kompozícióra cserélve jelentősen megnő az aminosavak kölcsönhatása (Mészáros et al., 2009). Így működik az ANCHOR-algoritmus. A becsült kötőhelyek jól egybeesnek a röntgenesek által meghatározott kötőhelyekkel, amelyeket az atomi kontaktusok számából becsültek (Dosztányi et al., 2009). Ezt sokféle fehérjén ellenőriztük, protomereken és oligomereken is. Az irodalomban egyik leggyakrabban emlegetett részben rendezetlen fehérje az emberi p53 tumorrepresszor fehérje. Ez egy négy azonos alegységből álló tetramer, amely a rendezetlen részein számos más fehérjével alakít ki funkcionálisan releváns kölcsönhatást. A röntgendiffrakciós vizsgálatok szerint a legnagyobb elektronsűrűségű, tehát a legmarkánsabb kötőhely ott van, ahol a négy alegység egymáshoz kötődik. Itt kapjuk a legerősebb jelet az ANCHOR-algoritmussal is. Pedig az ANCHOR szempontjából ez egy kakukktójas. Itt ugyanis nem egy más aminosav-összetételű globuláris fehérjéhez történik a kapcsolódás, hanem azonos kompozíciójú szakaszok között lép fel a csatolt kötődés és rendeződés (Fichó et al., 2016).

Jelenleg ezzel az eredetileg szabályt erősítő kivételnek tűnő jelenséggel, a rendezetlen fehérjék egymás közötti kölcsönhatásával foglalkozunk. Fiatal munkatársaim, *Mészáros Bálint* és *Fichó Erzsébet* átszitalták a fehérje-térszerkezeti

adatbankot, a PDB-t és számos más adatbázis, valamint cikk felhasználásával létrehoztak egy új, MFIB nevű adatbázist. Ebben az adatbázisban ilyen komplexekből mintegy kétszáz található. Megkezdjük az adatok elemzését. Szeretnénk feltárni ezeknek a kölcsönhatásoknak a fizikai hátterét és ha ez sikerül, a fizikai háttér felhasználásával szerkezetbecslő programot készíteni. Mivel a rendezetlen fehérjék, illetve fehérjerészek többsége a biokémiai folyamatok szabályozásában vesz részt, azt várjuk, hogy egyebek között megtudjuk, hogyan regulálódnak a reguláló fehérjék. Sőt, információt nyerhetünk arról is, mi történik a globuláris fehérjék szerkezetének szerveződésekor, amikor a rendezetlen láncrészek kölcsönhatásával rendezett fehérjeszerkezet alakul ki. Ezzel vissza is térünk a közel fél évszázados munkásságom alapkérdéséhez, a fehérjeszerkezet szerveződési alapelveinek feltárásához, de most már ehhez egy sokkal nagyobb, komplexebb fehérjevilágban tudunk dolgozni. A komplexitás növekedése nem nehezíti, hanem segíti a fehérjeszerkezet szerveződésének megértését. A rendezetlen fehérjék részvételével megvalósuló különböző típusú csatolt kötődés és szerkezetstabilizálódás fizikai hátterének feltárása segíti a rendezett fehérjék szerkezetének kialakulása során fellépő, de vizsgálhatatlanul gyors folyamatok megértését is. Remélem, hogy belátható időn belül ezekről az eredményekről is beszámolhatok.

Befejezésül köszönetemet fejezem ki egykori közvetlen előljáróimnak, néhai Elődi Pálnak, továbbá Harold A. Scheragának, Clare Woodwardnak és Závodszky Péternek a rengeteg segítségért, biztatásért és a velem való foglalkozáshoz elengedhetetlen türelemért, továbbá Falus András, Hudecz Ferenc, Patthy László és Penke Botond rendes tagoknak, akik levelező tagságra ajánlottak, támogató érvelésükért. Köszönet illeti egykori és jelenlegi diákjaimat, kollégáimat a gyakran egészen kiváló, ezer körüli hivatkozást eredményező munkáikért. Szeretném hangsúlyozni, hogy a több mint húsz éven át csoportomban dolgozó Dosztányi Zsuzsa és Tusnády Gábor, jelenlegi Lendület csoportvezetők nélkül ez az életmű nem valósulhatott volna meg. Ugyancsak köszönöm körülbelül százharminc cikkem további, nagyjából szintén százharminc társszerzőjének az együttműködését is. Közülük kiemelem Tompa Pétert, akivel bár soha nem dolgoztunk egy csoportban, 2001 és 2009 között tizenhat cikkünk jelent meg. Végül, de nem utolsósorban, köszönöm feleségemnek, Évinek, hogy fél évszázada velem van és szeretetével tüntet ki. Egyetemi vizsgáimtól az akadémiai székfoglalóra készülődésig végigizgulta értem ezt az ötven évet.

IRODALOM

- Cserző M. – Eisenhaber, F. – Eisenhaber, B. – Simon I. (2002): On Filtering False Positive Transmembrane Protein Predictions. *Protein Engineering*, 15, 9, 745–752. DOI: 10.1093/protein/15.9.745 <https://goo.gl/96HW9W>
- Cserző M. – Eisenhaber, F. – Eisenhaber, B. – Simon I. (2004): TM or Not TM: Transmembrane Protein Prediction with Low False Positive Rate Using DAS-TMfilter. *Bioinformatics*, 20, 1, 136–137. DOI: 10.1093/bioinformatics/btg394 <https://goo.gl/ZFbGJY>
- Cserző M. – Wallin, E. – Simon I. et al. (1997): Prediction of Transmembrane Alpha-Helices in Prokaryotic Membrane Proteins: The Dense Alignment Surface Method. *Protein Engineering*, 10, 6, 673–676. DOI: 10.1093/protein/10.6.673 <https://goo.gl/Mzz3ud>
- Dosztányi Zs. – Csizmok V. – Tompa P. – Simon I. (2005a): The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. *Journal of Molecular Biology*, 347, 4) 827–839. DOI: 10.1016/j.jmb.2005.01.071 <http://www.ttk.mta.hu/pdf/IUPred1.pdf>
- Dosztányi Zs. – Csizmok V. – Tompa P. – Simon I. (2005b): Iupred: Web Server for the Prediction of Intrinsically Unstructured Regions of Proteins Based on Estimated Energy Content. *Bioinformatics*, 21, 16, 3433–3434. DOI: 10.1093/bioinformatics/bti541 <https://academic.oup.com/bioinformatics/article-lookup/DOI/10.1093/bioinformatics/bti541>
- Dosztányi Zs. – Fiser A. – Simon I. (1997): Stabilization Centers in Proteins: Identification, Characterization and Predictions. *Journal of Molecular Biology*, 272, 4, 597–612. DOI: 10.1006/jmbi.1997.1242
- Dosztányi Zs. – Mészáros B. – Simon I. (2009): ANCHOR: Web Server For Predicting Protein Binding Regions in Disordered Proteins. *Bioinformatics*, 25, 20, 2745–2746. DOI: 10.1093/bioinformatics/btp518 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2759549/pdf/btp518.pdf>
- Fichó E. – Mészáros B. – Simon I. (2016): Two-state Protein Complexes. (poster) *F1000Research*, 5, 2147 DOI: 10.7490/f1000research.1112983.1 <https://f1000research.com/posters/5-2147>
- Fuxreiter M. – Tompa P. – Simon I. (2007): Local Structural Disorder Imparts Plasticity on Linear Motifs. *Bioinformatics*, 23, 8, 950–956. DOI: 10.1093/bioinformatics/btm035 <https://goo.gl/PvJAPF>
- Kozma D. – Simon I. – Tusnády G. E. (2013): PDBTM: Protein Data Bank of Transmembrane Proteins after 8 Years. *Nucleic Acids Research*, 41, D1, D524–D529. DOI: 10.1093/nar/gks1169 <https://academic.oup.com/nar/article-lookup/DOI/10.1093/nar/gks1169>
- Magyar Cs. – Gromiha, M. M. – Sávolgy Z. – Simon I. (2016): The Role of Stabilization Centers in Protein Thermal Stability. *Biochemical and Biophysical Research Communications*, 471, 1, 57–62. DOI: 10.1016/j.bbrc.2016.01.181 <http://www.sciencedirect.com/science/article/pii/S0006291X16301826>
- Mészáros B. – Simon I. – Dosztányi Z. (2009): Prediction of Protein Binding Region in Disordered Proteins. *PLOS Computational Biology*, 5, 5, Paper E1000376. <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000376>
- Simon I. (1971): Determination of Small Alternations in the Radius of Gyration by Small-Angle X-Ray Scattering. *Journal of Applied Crystallography*, 4, 317–318.
- Simon I. – Glasser, L. – Scheraga, H. A. (1991): Calculation of Protein Conformation as an Assembly of Stable Overlapping Segments: Application to Bovine Pancreatic Trypsin Inhibitor. *Proceedings of the National Academy of Sciences of the USA*, 88, 9, 3661–3665. DOI: 10.1073/pnas.88.9.3661 <http://www.pnas.org/content/88/9/3661.full.pdf>
- Simon I. – Glasser, L. – Scheraga, H. A. – Manley R. S. (1988b): Structure of Cellulose. 2. Low-Energy Crystalline Arrangements. *Macromolecules*, 21, 4, 990–998. DOI: 10.1021/ma00182a025

- Simon I. – Némethy G. – Scheraga, H. A. (1978): Conformational Energy Calculations of the Effect of Sequence Variation on the Conformation of Two Tetrapeptides. *Macromolecules*, 11, 4, 797–804. DOI: 10.1021/ma60064a035
- Simon I. – Scheraga, H. A. – Manley, R. S. (1988a): Structure of Cellulose. 1. Low-Energy Conformations of Single Chains. *Macromolecules*, 21, 4, 983–990. DOI: 10.1021/ma00182a024
- Simon I. – Tüchsen, E. – Woodward, C. (1984): Effect of Trypsin Binding on the Hydrogen Exchange Kinetics of Bovine Pancreatic Trypsin Inhibitor Beta-Sheet Nhs. *Biochemistry*, 23, 9, 2064–2068. DOI: 10.1021/bi00304a028
- Tusnády G. E. – Dosztányi Z. – Simon I. (2004): Transmembrane Proteins in the Protein Data Bank: Identification and Classification. *Bioinformatics*, 20, 17, 2964–2972. DOI: 10.1093/bioinformatics/bth340 <https://goo.gl/t5spJS>
- Tusnády G. E. – Dosztányi Z. – Simon I. (2005a): TMDet: Web Server for Detecting Transmembrane Regions of Proteins by Using Their 3D Coordinates. *Bioinformatics*, 21, 7, 1276–1277. DOI: 10.1093/bioinformatics/bti121 <https://goo.gl/yiJiUb>
- Tusnády G. E. – Dosztányi, Z. – Simon, I. (2005b): PDB_TM: Selection and Membrane Localization of Transmembrane Proteins in the Protein Data Bank. *Nucleic Acids Research*, 33, 1, D275–D278. DOI: 10.1093/nar/gki002 <https://goo.gl/jVEzLv>
- Tusnády G. E. – Simon, I. (1998): Principles Governing Amino Acid Composition of Integral Membrane Proteins: Application To Topology Prediction. *Journal of Molecular Biology*, 283, 2, 489–506. DOI: 10.1006/jmbi.1998.2107
- Tusnády G. E. – Simon I. (2001): The HMMTOP Transmembrane Topology Prediction Server. *Bioinformatics*, 17, 9, 849–850. DOI: 10.1093/bioinformatics/17.9.849 https://www.researchgate.net/publication/11759946_The_HMMTOP_transmembrane_topology_prediction_server
- Tüdös É. – Cserző M. – Simon I. (1990): Predicting Isomorphic Residue Replacements for Protein Design. *International Journal of Peptide and Protein Research*, 36, 3, 236–239. DOI: 10.1111/j.1399-3011.1990.tb00973.x
- Woodward, C. – Simon I. – Tüchsen, E. (1982): Hydrogen Exchange and the Dynamic Structure of Proteins. *Molecular and Cellular Biochemistry*, 48, 3, 135–160. DOI: 10.1007/BF00421225 https://www.researchgate.net/publication/16440722_Hydrogen_Exchange_and_the_Dynamic_Structure_of_Proteins