

# ÚJ LINEÁRIS MOTÍVUMOK AZONOSÍTÁSA BIOINFORMATIKAI MÓDSZEREKKEL AZ LC8 FEHÉRJE KÖLCSÖNHATÁSI HÁLÓZATÁBAN

## IDENTIFICATION OF NEW LINEAR MOTIFS IN THE INTERACTION NETWORK OF LC8 USING BIOINFORMATIC METHODS

Hajdu-Soltész Borbála<sup>1</sup>, Pajkos Máttyás<sup>2</sup>, Szaniszló Tamás<sup>3</sup>, Erdős Gábor<sup>4</sup>, Dosztányi Zsuzsanna<sup>5</sup>

<sup>1</sup>PhD-hallgató, tudományos segédmunkatárs, akantusz@caesar.elte.hu

<sup>2</sup>PhD-hallgató, tudományos segédmunkatárs, matyaspajkos@caesar.elte.hu

<sup>3</sup>PhD-hallgató, tudományos segédmunkatárs, tszaniszlo@caesar.elte.hu

<sup>4</sup>PhD-hallgató, tudományos segédmunkatárs, gerdos@caesar.elte.hu

<sup>5</sup>PhD, tudományos főmunkatárs, dosztanyi@caesar.elte.hu

MTA–ELTE Lendület Bioinformatikai Kutatócsoport, ELTE Természettudományi Kar Biokémia Tanszék

### ÖSSZEFOGLALÓ

A viszonylag gyenge, gyorsan felbomló fehérje-fehérje kölcsönhatások alapvető szerepet játszanak számos sejtbeli folyamat finomhangolásában. Ezek a kapcsolatok gyakran kis, kompakt funkcionális modulok, úgynevezett rövid lineáris motívumok (short linear motifs, SLiM) által valósulnak meg. A SLiM-ek kis méretük és nagyfokú flexibilitásuk révén változatos molekuláris funkciókat tudnak ellátni: részt vesznek a fehérjék célba juttatásában, szabályozzák a fehérjék aktivitását és kötődési tulajdonságait, valamint elősegítik a makromolekuláris komplexek összeszerelődését. Ugyanakkor speciális tulajdonságaik nagyban megnehezítik kísérletes azonosításukat. Éppen ezért a SLiM-mediált kölcsönhatások nagy részét jelenleg nem ismerjük. További lineáris motívumok és kölcsönhatásainak azonosításához újfajta bioinformatikai megközelítésekre van szükség. Jelen cikkünkben egy olyan új megközelítést mutatunk be, amely képes egy adott lineárismotívum-kötő doménhez új interakciós partnereket hatékonyan azonosítani. Módszerünket az LC8 dinein könnyűlánc fehérje példáján keresztül mutatjuk be. Az LC8 fehérjének már eddig is sok partnere volt ismert, de pontos feladata a mai napig nyitott kérdés. Módszerünk segítségével nemcsak új tagokkal bővítettük a partnerek listáját, de fény derült az LC8 egy korábban nem ismert szerepére is egy fontos, a szervek méretét szabályozó jelátviteli útvonalban is.

### ABSTRACT

In cells, weak and transient protein-protein interactions play critical roles in fine-tuning diverse biological processes. These interactions are commonly mediated by small, compact functional modules, called short linear motifs (SLiMs). Due to their small size and highly flexible nature, SLiMs can carry out diverse molecular functions and can be involved in targeting proteins to specific subcellular locations, regulating the activity and interactions of proteins and aiding the assembly of macromolecular complexes. However, their specific properties also make the

experimental discovery of SLiMs challenging. As a result, SLiM mediated interactions are still largely unexplored. In order to discover novel linear motif classes and their interactions, new bioinformatic approaches are needed. In this work, we introduce a novel bioinformatic method that enables the effective prediction of undiscovered binding partners for a given SLiM-binding domain. The approach was applied to the dynein light chain LC8, that already has several known binding partners but its actual function is still unclear. Our work not only presented novel LC8 binding partners, but also highlighted a new function of LC8 in the regulation of a highly important signaling pathway controlling organ size.

**Kulcsszavak:** bioinformatika, fehérje-fehérje kölcsönhatás, jelátvitel, rendezetlen fehérjék, lineáris motívumok

**Keywords:** bioinformatics, protein-protein interaction, signaling, disordered proteins, linear motifs

### LINEÁRIS MOTÍVUM-KÖLCSÖNHATÁSOK

A fehérjék jelentős része más fehérjékkel kialakított kölcsönhatások során látja el a feladatát. Ezeket az ún. fehérje-fehérje kölcsönhatásokat csoportosíthatjuk aszerint, hogy a részt vevő fehérjék milyen szerkezeti elemeikkel kapcsolódnak össze. A legrégebben ismert kölcsönhatási típus a domén-domén interakció, amely jellemzően kiterjedt felületen történő, erős kapcsolatok kialakulását eredményezi, és nagy méretű, stabil fehérjekomplexekben, valamint az enzim-inhibitor kölcsönhatásoknál fordul elő. Vannak azonban olyan funkciók, amelyek esetében előnyös, ha a kölcsönhatás viszonylag gyenge, és csak rövid ideig áll fenn. Ilyen átmeneti kapcsolatok kialakítására adnak lehetőséget a speciális globuláris doménekhez kapcsolódó lineáris motívumok (Van Roey et al., 2014). A lineáris motívumok (Short Linear Motifs, SLiM) a doménekkal ellentétben csak néhány (3–10) aminosav hosszúságúak, és leggyakrabban a fehérjék rendezetlen – vagyis jól meghatározható 3D-térszerkezettel nem rendelkező – régióiban helyezkednek el. Sokféle fehérje-fehérje kölcsönhatás közvetítésében játszanak szerepet, amelyek révén meghatározzák partnerük sejten belüli elhelyezkedését, befolyásolják azok működését, aktivitását és élettartamát. A lineáris motívumok lehetőséget adnak arra, hogy az adott fehérje sokféle, különböző szerkezetű és funkciójú partnerrel is kapcsolatba kerülhessen. Ezen adottságoknak köszönhetően a lineáris motívumok, illetve az ezeket hordozó fehérjék a szabályozási és jelátviteli útvonalak központi elemei.

A SLiM-ek alapvetően eukarióta fehérjékben fordulnak elő, de vírusokban és baktériumokban is írtak le motívum kölcsönhatásokat. A lineáris motívumok legnagyobb gyűjteménye az ELM- (Eukaryotic Linear Motif) adatbázis (Dinkel et al., 2016), amelyben jelenleg 262 féle motívumosztályt találunk, melyekhez összesen 197 fajban 3030 kísérletesen igazolt lineáris motívum tartozik. Bár az

adatbázis időről időre bővül az újabb eredményekkel, még mindig messze járunk attól, hogy az összes előfordulást ismerjük: a funkcionális lineáris motívumok valószínű száma milliós nagyságrendűre becsülhető (Tomba et al., 2014). Ez a hatalmas szakadék a jelenleg ismert és még felfedezésre váró SLiM-kölcsönhatások száma között arra vezethető vissza, hogy a lineáris motívum-kölcsönhatások mind kísérletesen, mind számítógépes módszerekkel nehezebben vizsgálhatóak az egyéb kölcsönhatásokhoz képest, és feltérképezésükre csak az utóbbi időben jelentek meg nagyskalás kísérleti megközelítések (Blikstad–Ivarsson, 2015).

A SLiM-ek evolúciós szempontból is eltérően viselkednek a globuláris doménekhez képest. A domének általában nagymértékű konzerváltságot mutatnak a rokon fehérjék között, kialakulásuk fő hajtóereje a genomduplikáció és a rekombináció. Ezzel szemben a lineáris motívumokra nagyfokú evolúciós plaszticitás jellemző: egy adott funkciót ellátó motívum több különböző fehérjében, egymástól függetlenül, konvergens evolúció révén is kialakulhat. Ezen tulajdonságok alapján állították fel az *ex-nihilo* SLiM evolúciós elméletet, amely szerint egy új SLiM létrejöhet (vagy akár meg is szűnhet) random mutációk által (Davey et al., 2015). Az így kialakult motívum ezután a kötőpartner felszínével együtt evolválódik, pozitív és negatív szelekciós mechanizmusok irányításával finomhangolva a kölcsönhatást. Az *ex-nihilo* motívumszületés egyik érdekes példája a SHOC2 gén által kódolt fehérje (Leucine-rich repeat protein SHOC-2), amelyben egy meghatározott aminosav cseréje egy nem funkcionális régió belül – gyakorlatilag a „semmiből” – alakít ki egy funkcionális motívumot, melyen keresztül a fehérjéhez egy lipidmolekula kötődik. Ez a mutáció a fehérje hibás működését eredményezi, mely hiba végső soron a Noonan-szindrómával hozható összefüggésbe. A lineáris motívumokat érintő mutációk számos egyéb betegség, többek között a rák kialakulásában is szerepet játszanak (Van Roey et al., 2014).

#### BIOINFORMATIKAI MEGKÖZELÍTÉSEK A LINEÁRIS MOTÍVUM-KÖLCSÖNHATÁSOK AZONOSÍTÁSÁRA

Egy adott lineáris motívum-kötő domén kölcsönhatási partnereinek feltérképezésében nagyon fontos szerepet játszanak a bioinformatikai módszerek. A lineáris motívumokat jellemezhetjük reguláris kifejezéssel (*1/a ábra*), szekvencialogóval (*1/b ábra*) vagy pozícióspecifikus pontozómátrixszal. Mindhárom módszer a kölcsönhatás szempontjából kritikus pozíciókban előforduló aminosavak leírását szolgálja. A pontozómátrix megadja, hogy a motívum egyes pozícióiban mekkora valószínűséggel fordul elő egy bizonyos aminosav a véletlen előforduláshoz képest. Ez a módszer alkalmas a szigorú definíciót adó reguláris kifejezéstől eltérő motívumok leírására is. Az új partnerek azonosítása általában azon az elven alapul, hogy egy adott doménhez kötődő fehérjékben hasonló motívumok megjelenése várható. Fontos azonban, hogy a lineáris motívum megléte a szekvenciában önmagában még

nem bizonyíték arra, hogy egy fehérje valódi kölcsönhatási partner. Mivel a kötő régió csak néhány aminosavból áll, motívumtalálatok véletlenül is megjelenhetnek, a humán proteomot tekintve akár milliós nagyságrendben is. Ezért a lineáris motívumon alapuló kölcsönhatási partnerek azonosításában a legnagyobb nehézséget az jelenti, hogy a valódi kötőrégiókat kiszűrjük a nagyszámú véletlen találat közül.

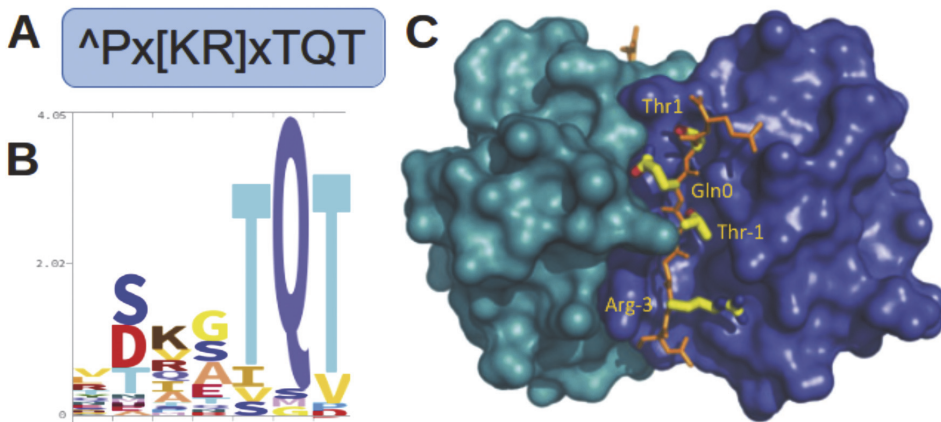
A biológiai releváns kölcsönhatások kiszűréséhez számos olyan bioinformatikai módszer nyújt segítséget, amelyek a fehérjék különböző tulajdonságait képesek megbecsülni pusztán az aminosav-szekvenciák alapján. Ilyen becsülhető tulajdonság például a fehérjék sejten belüli lokalizációja, mely alapján eldönthető, hogy az általunk vizsgált két fehérje valóban találkozhat-e a sejtben. Egy másik fontos szempont, hogy a kötőrégióknak a partner számára hozzáférhetőnek kell lennie, tehát a motívum nagyobb valószínűséggel található a fehérje rendezetlen részén, mintsem egy globuláris domén belsejében eltemetve. Ezekre a tulajdonságokra fehérje-térszerkezetekkel és szekvenciacsaldókkal való összehasonlítás, illetve rendezetlenségjelölés alapján lehet következtetni. A valódi találatok kiszűréséhez gyakran használt egyéb tulajdonságok lehetnek például a rendezetlen kötőrégiók és másodlagos szerkezeti elemek, illetve egyéb szerkezeti és funkcionális tulajdonságok. Végül, de nem utolsósorban, a SLiM-ek konzerváltságát is érdemes figyelembe venni. A rendezetlen régiók konzerváltsági profilja ugyanis egy speciális, SLiM-re jellemző mintázatot mutat a rokon fehérjék szekvenciaillesztésében. Ennek során a motívumok szigetszerűen emelkednek ki környezetükhöz képest, mivel ez a környezet sem szerkezeti, sem funkcionális megkötéssel nem rendelkezik, így általában szignifikáns konzerváltságot nem mutat. A szigetszerű konzerváltság alapján több olyan példát is leírtak már, amelyekben rendezetlen régiók egy-egy motívuma a fajok széles spektrumában mutatott szigetszerű konzerváltságot, és az így azonosított motívumok olyan alapvető biológiai funkciókban vettek részt, mint például a DNS-replikáció vagy sejtciklus-szabályozás.

Bár az egyes bioinformatikai módszerek hatékonysága eltérő, ezek kombinációja nagymértékben elősegíti a biológiai relevanciával bíró lineárismotívum-találatok azonosítását (Krystkowiak–Davey, 2017). Ezáltal vizsgálhatóak egyedi lineárismotívum-kötő rendszerek is, ami nemcsak a kölcsönhatási hálózat kiterjesztését eredményezi, de segíthet a meglévő módszerek továbbfejlesztésében is.

### AZ LC8 DINEIN KÖNNYŰLÁNC KÖLCSÖNHATÁSI HÁLÓZATA

Egy sokak által tanulmányozott, de még mindig nem teljesen feltérképezett, lineáris motívumon alapuló rendszer az LC8 fehérje interakciós hálózata (Rapali et al., 2011). Az LC8 a dinein könnyűláncok családjába tartozik, és először a dinein motorkomplex tagjaként írták le. A dinein motorkomplex a sejtben belüli transzportban vesz részt, a mikrotubuláris rendszer részeként vezikulákat és fehérje-

komplexekeket szállít. Kezdetben úgy vélték, hogy az LC8 funkciója a szállítandó anyagok specifikus felismerése és a motorkomplexhez való kapcsolása. Újabb kutatások azonban rávilágítottak arra, hogy az LC8 számos, a dinein komplextől független funkcióval is bír. Részt vesz többek között különböző jelátviteli útvonalakban, valamint a sejtsztódás és az apoptózis szabályozásában. Az LC8-at ezért ma már központi (hub) fehérjeként tartják számon, amely más fehérjék dimerizációjában és a dimerek stabilizálásában vesz részt. Az LC8 maga is homodimerként látja el feladatát. A dimerizáció során két párhuzamos kötőárok jön létre, ahová a kötőpartnerek rendezetlen régióiban található lineáris motívumok kötődnek. A kapcsolódás során a motívumok rendezetté válnak, és  $\beta$ -szál konformációt vesznek fel (1/c ábra). Jelenleg mintegy hatvan LC8-kötőmotívumot tartalmazó, különböző biológiai folyamatokban részt vevő partnert ismerünk, melyek nagy része „TQT-”motívumot tartalmaz (1/a ábra). Az LC8 fehérje általános funkciója alapján feltételezhető, hogy számos további partnere van még, melyek azonosítása fontos lehet a különböző biológiai folyamatok jobb megértésében.



**1. ábra.** Az LC8 fehérje kölcsönhatási partnereiben található motívumok leírása.

**A:** Reguláris kifejezés. A hét aminosavból álló szekvencia első helyén nem fordulhat elő prolin (P); a 2. és 4. pozícióban található „x” tetszőleges aminosavat jelöl; a 3. pozícióban lizin (K) vagy arginin (R) található; az utolsó három helyen pedig csak treonin (T), glutamin (Q) és treonin (T) fordulhat elő. **B:** Szekvencialógó. Az egyes pozíciókban a betűk mérete arányos az aminosav előfordulási valószínűségével. **C:** Két LC8 monomer (különböző árnyalatú szürkével jelölve) által kialakított kötősebben az EML3 fehérje kanonikus LC8-kötőmotívuma található (PDB kód: 2XQQ).

Munkánk során azt a célt tűztük ki, hogy további LC8-kötőpartnereket azonosítsunk. Irodalmi adatokat gyűjtve létrehoztunk egy adatbázist, amely a korábban kísérletesen igazolt LC8 partnerek kötőmotívumait tartalmazta. Ezeket felhasznál-

nálva elkészítettünk egy pozícióspecifikus pontozómátrixot, és összegyűjtöttük a lehetséges motívumokat humán proteomban, melyek azonban még mindig nagy számban tartalmaztak fals pozitív találatokat. Az ismert LC8 partnerek tulajdonságai alapján előállítottunk egy olyan szűrési kritériumrendszert, amely a sok millió találat közül nagy megbízhatósággal képes kiszűrni a valódi LC8-kötőmotívumokat. Ehhez létrehoztunk egy döntési fát, és bevezettünk egy új mérőszámot, ami alapján meg tudtuk határozni az optimális szűrési tulajdonságokat erre a rendszerre. A kapott kritériumok alapján azonosítottuk azokat a sejt belsőjében előforduló, nem globuláris doménben található, a fehérjék rendezetlen részein elhelyezkedő motívumokat, amelyek a pozícióspecifikus pontozómátrix alapján kedvező értékeket kaptak. Érdekes módon számos más tulajdonság az ismert LC8-motívumoknak csak egy részhalmazára teljesült, és nem tudtuk szűrési kritériumként felhasználni. Ilyenek voltak a motívumon belüli és annak környezetében található szerkezeti elemek vagy a rendezetlen kötőrégió megléte. A motívumkonzerváltság viszont fontos szűrési kritérium volt, amely az evolúciósan közeli fajokra alkalmazott konzerváltsági tulajdonságra épült.

Vizsgálataink eredményeképpen 72 új potenciális humán kötőpartnert azonosítottunk, melyek alapján az LC8 egy eddig nem ismert szerepére is fény derült (Erdős et al., 2017). Az eredményeink alapján az LC8 kapcsolatba lép az ún. Hippo jelátviteli útvonal két elemével: az angiomotin és a WWC-fehérjecsalád tagjaival. A Hippo-útvonal a szomszédos sejtek általi kontakt gátlás következtében, valamint a sejteket érő mechanikai hatásokra aktiválódik. Működésével gátolja a sejtosztódást elősegítő és a programozott sejthalált gátló gének átírását szabályozó fehérjék működését, ezáltal kontrollálja a sejtek és szervek növekedését (Yu–Guan, 2013). Az angiomotin és WWC-fehérjék közös tulajdonsága, hogy állványfehérjeként (scaffold) működve számos fehérjével alakítanak ki kölcsönhatásokat SLiM-eken keresztül, és interakciós platformot biztosítanak a jelátvitelt elindító és szabályozó fehérjéknek. Emellett az angiomotin fehérjecsalád tagjai végrehajtó (effektor) funkciót is betöltenek azáltal, hogy megkötik a sejtosztódás folytatásának szabályozásában részt vevő fehérjekomplexek egyes tagjait. A WWC-család tagjaiban kettő, az angiomotin családban egy LC8-kötőmotívumot azonosítottunk, amelyeket kísérletesen is igazoltunk (Erdős et al., 2017). Evolúciós vizsgálataink rámutattak arra, hogy a WWC-család fehérjeinek két kötőmotívuma erős konzerváltságot mutat az eukarióta fajok széles körében, ezzel szemben az angiomotin fehérjecsalád motívuma csak a gerincesek közös ősig követhető vissza. Ezek az eredmények alátámasztják az LC8 fehérjével való kölcsönhatás fontosságát, de rámutatnak arra is, hogy míg a WWC-fehérjék a Hippo-útvonalban univerzális szabályozóként vesznek részt az eukarióta fajok többségében, addig az angiomotinok megjelenése valószínűleg csak későbbi, gerinces „találmány” lehet. Az LC8-fehérje Hippo-útvonalban betöltött szerepe ezeknek az eredményeknek a tükrében sem teljesen tisztázott még. Valószínűsít-

hető azonban, hogy az angiominin és WWC-fehérjék dimerizációjáért, és a dimerek stabilizálásáért felelős, ezáltal egy új, eddig nem ismert finomhangolási lehetőséget teremthet a Hippo jelátviteli útvonalban.

## KONKLÚZIÓ

A lineáris motívumok alapvető biológiai folyamatok szabályozásában vesznek részt, de kölcsönhatásaik még nagyrészt feltáratlanok. Az LC8 lineárismotívum-kötő fehérje példáján keresztül bemutattuk, hogyan lehet bioinformatikai megközelítésből kiindulva új kötőmotívumokat azonosítani, amelyek nemcsak a kölcsönhatási hálózatot bővítik, de további vizsgálatokra érdemes, izgalmas új funkciót is felvetnek. Bár a pontos szűrési kritériumok részben specifikusak erre a rendszerre, a kidolgozott protokoll használható tetszőleges lineárismotívum-kötő domén esetén. Ennek kulcsa a döntési fa és a hozzá kapcsolódó mérőszám, amely lehetővé teszi az optimális szűrési feltételek meghatározását. Eredményeink rámutattak arra is, hogy a motívumok konzerváltsága rendkívül informatív a biológiailag releváns kölcsönhatások azonosítására. Ezt technikai okok miatt nehéz kiaknázni, ezért nagy szükség van a jelenlegi megközelítésnél sokkal hatékonyabban működő módszerekre. A nagyskálás kísérleteket kiegészítve a bioinformatikai vizsgálatok nagyban hozzájárulhatnak a lineárismotívum-rendszerek feltérképezéséhez, és ezáltal annak megértéséhez, hogy ezek a viszonylag gyenge, tranziens kölcsönhatások hogyan járulnak hozzá a sejten belüli dinamikus szabályozó folyamatok komplex rendszeréhez.

## IRODALOM

- Blikstad, C. – Ivarsson, Y. (2015): High-Throughput Methods for Identification of Protein-Protein Interactions Involving Short Linear Motifs. *Cell Communication and Signaling*, 13, 38. DOI: 10.1186/s12964-015-0116-8, <https://biosignaling.biomedcentral.com/articles/10.1186/s12964-015-0116-8>
- Davey, N. E. – Cyert, M. S. – Moses, A. M. (2015): Short Linear Motifs – Ex Nihilo Evolution of Protein Regulation. *Cell Communication and Signaling*, 13, 43. DOI: 10.1186/s12964-015-0120-z, <https://biosignaling.biomedcentral.com/articles/10.1186/s12964-015-0120-z>
- Dinkel, H. – Van Roey, K. – Michael S. et al. (2016): ELM 2016 – Data Update and New Functionality of the Eukaryotic Linear Motif Resource. *Nucleic Acids Research*, 44, D294–300. DOI: 10.1093/nar/gkv1291, <https://academic.oup.com/nar/article/44/D1/D294/2503097>
- Erdős G. – Szaniszló T. – Pajkos M. et al. (2017): Novel Linear Motif Filtering Protocol Reveals the Role of the LC8 Dynein Light Chain in the Hippo Pathway. *PLoS Computational Biology*, 13, 12. DOI: 10.1371/journal.pcbi.1005885, <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005885>

- Krystkowiak, I. – Davey, N. E. (2017): SLiMSearch: A Framework for Proteome-Wide Discovery and Annotation of Functional Modules in Intrinsically Disordered Regions. *Nucleic Acids Research*, 45, W464–W469. DOI: 10.1093/nar/gkx238, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5570202/>
- Rapali P. – Szenes Á. – Radnai L. et al. (2011): DYNLL/LC8: A Light Chain Subunit of the Dynein Motor Complex and beyond. *The FEBS Journal*, 278, 17, 2980–2996. DOI: 10.1111/j.1742-4658.2011.08254.x, <http://onlinelibrary.wiley.com/doi/10.1111/j.1742-4658.2011.08254.x/full>
- Tompa P. – Davey, N. E. – Gibson, T. J. et al. (2014): A Million Peptide Motifs for the Molecular Biologist. *Molecular Cell*, 55, 2, 161–69. DOI: 10.1016/j.molcel.2014.05.032, [http://www.cell.com/molecular-cell/fulltext/S1097-2765\(14\)00562-0](http://www.cell.com/molecular-cell/fulltext/S1097-2765(14)00562-0)
- Van Roey, K. – Uyar, B. – Weatheritt, R. J. et al. (2014): Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chemical Reviews*, 114, 13, 6733–6778. DOI: 10.1021/cr400585q
- Yu, F.-X. – Guan, K.-L. (2013): The Hippo Pathway: Regulators and Regulations. *Genes & Development*, 27, 4, 355–371. DOI: 10.1101/gad.210773.112, <http://genesdev.cshlp.org/content/27/4/355.long>