



Data Article

Data on evolution of intrinsically disordered regions of the human kinome and contribution of FAK1 IDRs to cytoskeletal remodeling



Jaymin J. Kathiriya^a, Ravi Ramesh Pathak^a,
Alexandr Bezginov^c, Bin Xue^d, Vladimir N. Uversky^{e,f,g},
Elisabeth R.M. Tillier^c, Vrushank Davé^{a,b,*}

^a Morsani College of Medicine, Department of Pathology and Cell Biology, University of South Florida, Tampa, FL 33612, USA

^b Department of Cancer Biology and Evolution, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA

^c Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

^d Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, Tampa, FL 33620, USA

^e Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA

^f USF Health Byrd Alzheimer's Research Institute, University of South Florida, Tampa, FL 33612, USA

^g Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russian Federation

ARTICLE INFO

Article history:

Received 4 October 2016

Received in revised form

21 November 2016

Accepted 30 November 2016

Available online 8 December 2016

Keywords:

Focal adhesion Kinase-1

Protein kinases

Evolution of intrinsically disordered regions

Cytoskeletal remodeling

ABSTRACT

We present data on the evolution of intrinsically disordered regions (IDRs) taking into account the entire human protein kinome. The evolutionary data of the IDRs with respect to the kinase domains (KDs) and kinases as a whole protein (WP) are reported. Further, we have reported its post translational modifications of FAK1 IDRs and their contribution to the cytoskeletal remodeling. We also report the data to build a protein-protein interaction (PPI) network of primary and secondary FAK1-interacting hybrid proteins. Detailed analysis of the data and its effect on FAK1-related functions have been described in

DOI of original article: <http://dx.doi.org/10.1016/j.bbapap.2016.10.002>

* Corresponding author. Tel.: 813 974 0930.

E-mail address: vdave@health.usf.edu (V. Davé).

<http://dx.doi.org/10.1016/j.dib.2016.11.099>

2352-3409/© 2017 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

“Structural pliability adjacent to the kinase domain highlights contribution of FAK1 IDRs to cytoskeletal remodeling” (Kathiriya et. al., 2016) [1].

© 2017 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Biology
More specific sub- ject area	Bioinformatics, Computational and Systems Biology, Evolutionary Biology
Type of data	Tables and Figures
How data was acquired	Computational Analysis
Data format	Raw and Filtered
Experimental factors	None
Experimental features	Intrinsically disordered regions of the human protein kinome have been com- putationally analyzed.
Data source location	N/A
Data accessibility	With this article

Value of the data

- This data includes detailed information on evolution of each of the IDRs in the entire human protein kinome, which can be leveraged to study either individual IDRs within a kinase or a group of IDRs within a kinase family.
- Protein-protein interaction (PPI) data reported here can be utilized to further explore the contribution of IDRs in facilitating kinase functions.
- Our data on the utility of IDRs by FAK1 to relay its cytoskeleton-related signaling can lead to future experiments to determine the role of IDRs in the entire process of cytoskeletal remodeling.
- KD-adjacent IDRs, which are under higher evolutionary pressure, can be further analyzed to identify functionally important regions within these IDRs for future therapeutic targeting of kinases.

1. Data

Data reported here are related to the article entitled “Structural Pliability Adjacent to the Kinase Domain Highlights Contribution of FAK1 IDRs to Cytoskeletal Remodeling” [1]. Six figures and nine tables are presented in this article. The figures illustrate the function of IDRs in FAK1 and its effects on cytoskeletal remodeling. The tables provide raw data utilized to build PPI networks. Evolutionary scores of IDRs, kinase domains, and whole kinases are also reported in tables. One single Microsoft Excel file is provided with one table on each of the nine sheets (Fig. 6).

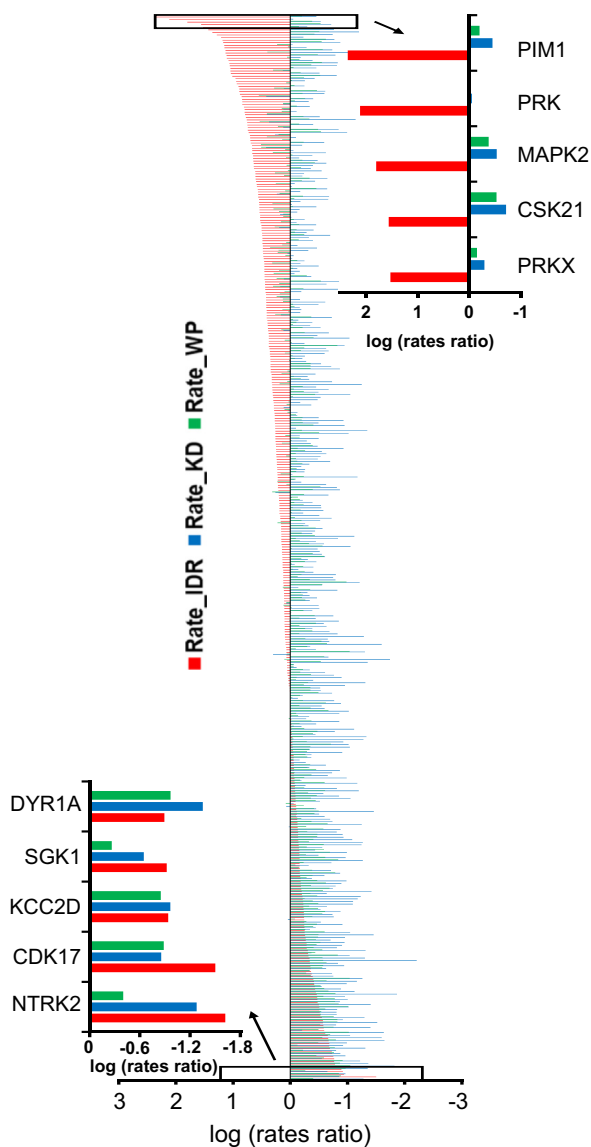


Fig. 1. Evolution of IDRs, KDs, and whole kinases (WP): The rates of evolution are depicted in the bar graph for each kinase: WP (green), its KD (blue), and average of all the IDRs (red) for each kinase. A total of 414 kinases are depicted after they were sorted on the basis of the lowest to highest evolutionary rates of their IDRs. For brevity, we have also included the bottom inset depicting five kinases with lowest evolutionary rates of their IDRs. Likewise, the top inset depicts five kinases with highest evolutionary rates of their IDRs. The boxes in the main image depict the location of the top and bottom insets.

2. Experimental design, materials and methods

2.1. Disorderliness in the kinome

We predicted intrinsic disorder in the human kinome using Ponder-FIT software [2] (Figs. 1 and 2A; Supplementary Table 1). PONDR-FIT is an artificial neural network-aided meta-predictor of

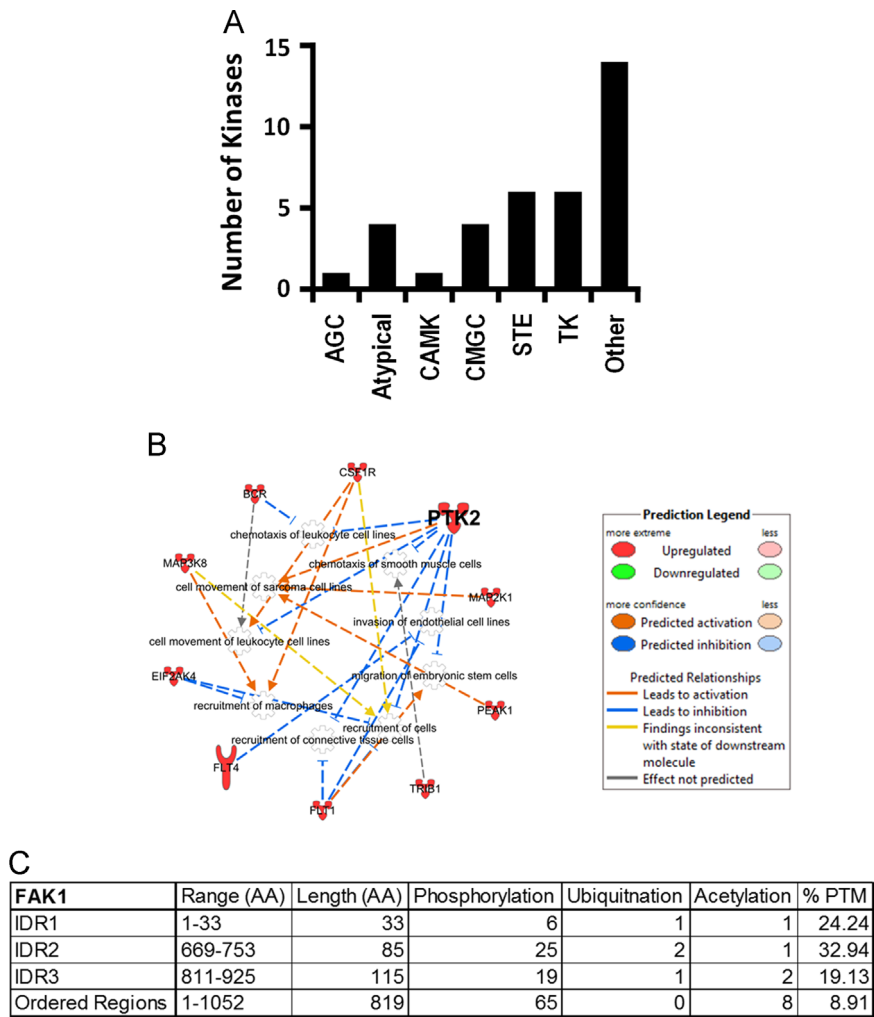


Fig. 2. Identification of 36 Kinases with IDRs extending into the KDs and cellular migration as a significantly enriched molecular function. (A) 36 kinases were identified which had their IDRs extending into the KDs. Here, their distribution across seven different kinase groups is depicted. Raw data are provided in [Supplementary Tables 4 and 5](#). (B) We identified cellular migration as a significantly enriched molecular function via the set of 36 kinases when compared to rest of the kinome ([Fig. 4C](#) in Ref. [1]). Network of proteins involved in cellular migration is depicted with predicted inter-relationships. Red color-filled molecules are part of 36 kinases. (C) Post translation modifications (PTM) were predicted in IDR₁ (flanking the KD), IDR₂ (2nd closest IDR to the KD), IDR₃ (N-terminal region, which is farthest away from the KD), and rest of the ordered regions of FAK1. These PTMs include phosphorylation, acetylation, and ubiquitination. % PTM was calculated as described in methods.

disordered residues. PONDR-FIT combines output of 7 different individual disorder predictors to increase confidence of disorder prediction by an average of 11% as compared to individual disorder predictors [2]. PONDR-FIT utilizes the following amino acid characteristics to predict disorder residues: Amino composition, amino acid sequence complexity, amino acid position specific scoring matrices, hydrophobicity and net charge of amino acid sequence, and pairwise interaction energy between amino acids of a given protein We considered the residues with disorder scores of ≥ 0.5 to have structure breaking propensities, or as we call it intrinsically disordered residues, as previously described [3]. A long disordered region with a stretch of at least 25 such amino acids

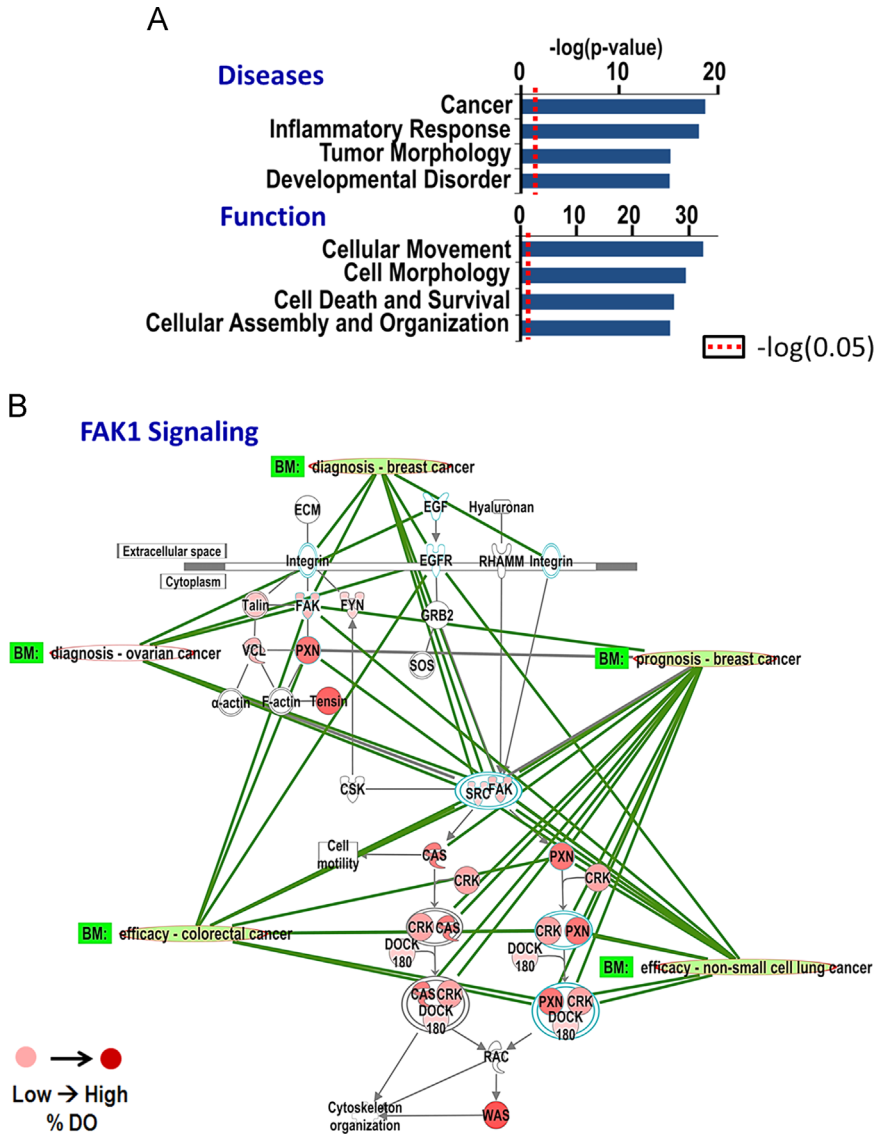


Fig. 3. FAK1 signaling is mediated by hybrid proteins. (A) Functional analysis of the 98 proteins revealed that FAK1 relays its functions of modulating cellular movement, morphology and organization through its primary interactome, which enriches important diseases such as cancer. Analysis was performed using IPA. Red dotted line represents the threshold for significance ($p < 0.05$). (B) FAK1 signaling pathway was enriched by the FAK1 primary interactome. Overlaying percent disorderliness of hybrid proteins revealed that cytoskeleton organization pathways are modulated by FAK1 interacting protein partners, most being intrinsically disordered. Intensity of red color denotes percent disorder in a protein in this pathway. The signaling pathway was further analyzed to identify proteins that serve as clinical markers for various cancer types. Intrinsically disordered hybrid proteins were identified as important biomarkers for a number of cancers.

constituted an IDR in our analysis. Previously reported disorder prediction of 504 kinases was used to calculate the fraction of total disordered amino acids in each kinase [4]. An amino acid labeled with disorder score of 0.5 or greater was considered as contributing to protein disorder. The total number of disordered amino acids was divided by the total number of amino acids present in a given kinase to calculate % DO (percent disorder) in the kinome. KINOMERender [5], a visualization tool for overlapping annotations on a phylogenetic tree of protein kinases annotated the kinome dendrogram with % DO for each of the 504 kinases. We have excluded proteins without confirmed kinase domains in UniProt database. Kinome dendrogram illustration was reproduced, courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com). Intrinsic disorder prediction of FAK1 and its orthologs was performed using Ponder-FIT [2], IUPred-L [6], IUPred-S [6], VSL2 [7], VSL3 [3], VLXT [8], Espritz [9], PrDOS [10].

2.2. Evolutionary analysis of kinases, KDs, and IDRs

The relative rates of evolution for the proteins and their domains (Figs. 1 and 2; Supplementary Tables 1–3) were calculated as described by Kathiriya et al. [1].

2.3. PTM analysis of IDRs

We predicted phosphorylation sites using disPhos [11] and netphos [12], acetylation sites using PAIL [13], and ubiquitylation sites using UbPred [14]. These PTM predictors, in addition to predicting novel PTM sites, also predict and report the PTM sites that are already experimentally validated. We have cross-validated the data to ensure that all the PTM data reported for FAK1 in UniProt are included in our PTM data generated using these softwares. We calculated the abundance of PTM sites and normalized it to the total length of ordered or disordered regions as per the following equation (Fig. 2C):

$$\% \text{ PTM} = \frac{\text{Phosphorylation Sites} + \text{Ubiquitination Sites} + \text{Acetylation Sites}}{\text{Length of the region (AA residues)}} \times 100$$

2.4. Network and functional analysis

Core analysis of 36 kinases was performed using Ingenuity Pathway Analysis (IPA) as described by Kathiriya et al. [15] (Fig. 3). Network of cellular migration as a significantly enriched function of the 36 kinases was identified. Disease and functional enrichment was performed as described by Kathiriya et al. [4].

2.5. Derivation of PPIs

Experimentally validated protein-protein interaction (PPI) data of the 36 kinases and that of FAK1 interacting proteins was assembled using manual data curation and various softwares including as described previously [4] (Supplementary Tables 5–8). PPI network was constructed and visualized using Cytoscape [16] (Fig. 4). Network analysis was performed to identify topologically significant hubs from the PPI networks using Network Analyzer [17] and CentiScaPe plug in tools [18] (Fig. 5). Further, canonical signaling pathways by IDR-interacting proteins of FAK1 interactome were enriched (Fig. 6).

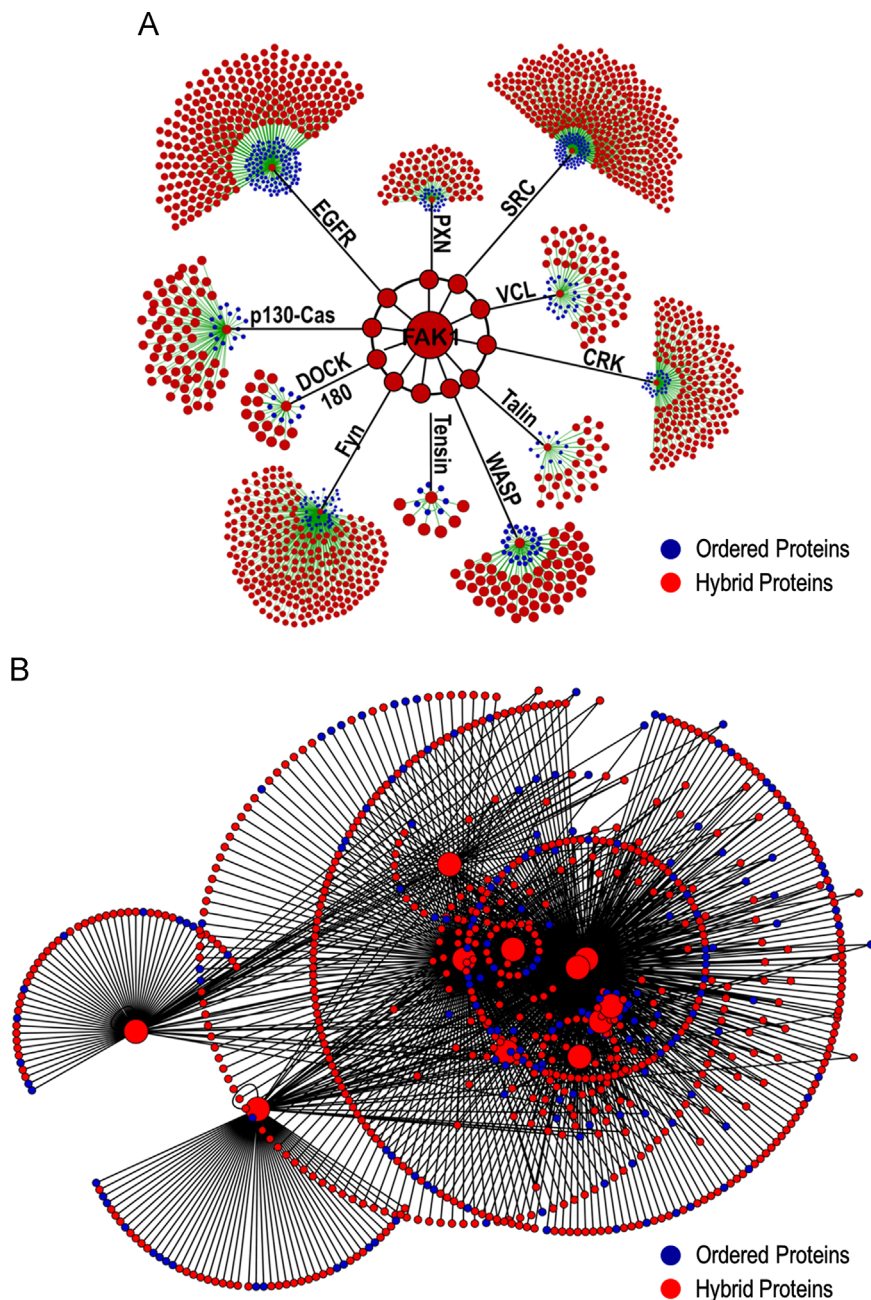


Fig. 4. Intrinsic disorder provides structural pliancy in cytoskeletal remodeling via FAK1 and its secondary interactome. (A) Primary interactomes derived from 11 hybrid FAK1 interacting cytoskeletal proteins formed the FAK1 secondary interactome. Our analysis identified that 77% of FAK1 secondary interactome is comprised of hybrid proteins. Blue circles represent structured proteins and red circles represent hybrid proteins. (B) Overlay of the 11 interactomes onto one another revealed a high degree of crosstalk. 30% of the proteins interacted with at least 2 of the 11 cytoskeletal hybrid proteins while 27 proteins interacted with at least 5 of the 11 cytoskeletal hybrid proteins, indicating that these 11 cytoskeletal hybrid proteins are major hub proteins governing the cytoskeletal remodeling network.



Fig. 6. Enrichment of canonical signaling pathways by IDR-interacting proteins of FAK1 interactome. Mapped interactome of FAK1 was divided into two interactomes. A – Ordered region-interacting proteins; B – IDR-interacting proteins. Differentially enriched canonical pathways were identified by using comparison analysis feature of IPA. Raw data is described in [Supplementary Table 9](#).

Transparency document. Supporting information

Transparency document associated with this article can be found in the online version at: <http://dx.doi.org/10.1016/j.dib.2016.11.099>.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.11.099>.

References

- [1] J.J. Kathiriya, et al., Structural pliability adjacent to the kinase domain highlights contribution of FAK1 IDRs to cytoskeletal remodeling, *Biochim Biophys. Acta* 1865 (2016) 43–54. <http://dx.doi.org/10.1016/j.bbapap.2016.10.002>.
- [2] B. Xue, R.L. Dunbrack, R.W. Williams, A.K. Dunker, V.N. Uversky, PONDR-FIT: a meta-predictor of intrinsically disordered amino acids, *Biochim Biophys. Acta* 1804 (2010) 996–1010. <http://dx.doi.org/10.1016/j.bbapap.2010.01.011>.
- [3] Z. Obradovic, et al., Predicting intrinsic disorder from amino acid sequence, *Proteins* 53 (Suppl 6) (2003) 566–572. <http://dx.doi.org/10.1002/prot.10532>.
- [4] J.J. Kathiriya, et al., Presence and utility of intrinsically disordered regions in kinases, *Mol. Biosyst.* (2014), <http://dx.doi.org/10.1039/c4mb00224e>.
- [5] M. Chartier, T. Chenard, J. Barker, R. Najmanovich, Kinome Render: a stand-alone and web-accessible tool to annotate the human protein kinome tree, *PeerJ* 1 (2013) e126. <http://dx.doi.org/10.7717/peerj.126>.
- [6] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins, *J. Mol. Biol.* 347 (2005) 827–839. <http://dx.doi.org/10.1016/j.jmb.2005.01.071>.
- [7] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, A.K. Dunker, Exploiting heterogeneous sequence properties improves prediction of protein disorder, *Proteins* 61 (Suppl 7) (2005) 176–182. <http://dx.doi.org/10.1002/prot.20735>.
- [8] P. Romero, et al., Sequence complexity of disordered protein, *Proteins* 42 (2001) 38–48.
- [9] I. Walsh, A.J. Martin, T. Di Domenico, S.C. Tosatto, ESpritz: accurate and fast prediction of protein disorder, *Bioinformatics* 28 (2012) 503–509. <http://dx.doi.org/10.1093/bioinformatics/btr682>.
- [10] T. Ishida, K. Kinoshita, PrDOS: prediction of disordered protein regions from amino acid sequence, *Nucleic acids Res.* 35 (2007) W460–W464. <http://dx.doi.org/10.1093/nar/gkm363>.
- [11] L.M. Iakoucheva, et al., The importance of intrinsic disorder for protein phosphorylation, *Nucleic acids Res.* 32 (2004) 1037–1049. <http://dx.doi.org/10.1093/nar/gkh253>.
- [12] N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, *J. Mol. Biol.* 294 (1999) 1351–1362. <http://dx.doi.org/10.1006/jmbi.1999.3310>.
- [13] A. Li, Y. Xue, C. Jin, M. Wang, X. Yao, Prediction of Nepsilon-acetylation on internal lysines implemented in Bayesian Discriminant Method, *Biochem. Biophys. Res. Commun.* 350 (2006) 818–824. <http://dx.doi.org/10.1016/j.bbrc.2006.08.199>.
- [14] P. Radivojac, et al., Identification, analysis, and prediction of protein ubiquitination sites, *Proteins* 78 (2010) 365–380. <http://dx.doi.org/10.1002/prot.22555>.
- [15] S. Muller, A. Chaikwad, N.S. Gray, S. Knapp, The ins and outs of selective kinase inhibitor development, *Nat. Chem. Biol.* 11 (2015) 818–821. <http://dx.doi.org/10.1038/nchembio.1938>.
- [16] M.E. Smoot, K. Ono, J. Ruscheinski, P.L. Wang, T. Ideker, Cytoscape 2.8: new features for data integration and network visualization, *Bioinformatics* 27 (2011) 431–432. <http://dx.doi.org/10.1093/bioinformatics/btq675>.
- [17] Y. Assenov, F. Ramirez, S.E. Schelhorn, T. Lengauer, M. Albrecht, Computing topological parameters of biological networks, *Bioinformatics* 24 (2008) 282–284. <http://dx.doi.org/10.1093/bioinformatics/btm554>.
- [18] G. Scardoni, M. Patterlini, C. Laudanna, Analyzing biological network parameters with CentiScaPe, *Bioinformatics* 25 (2009) 2857–2859. <http://dx.doi.org/10.1093/bioinformatics/btp517>.