

PSINDB: A comprehensive database of postsynaptic protein-protein interactions

Zsafia E. Kalman¹, Dániel Dudola¹, Bálint Mészáros², Zoltán Gáspári¹, Laszlo Dobson³

¹ Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter u. 50/A, 1083 Budapest, Hungary

² Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstraße 1, Heidelberg, 69117, Germany

³ Institute of Enzymology, Research Centre for Natural Sciences, Magyar Tudósok Körútja 2, Budapest, 1117, Hungary

Abstract

The postsynaptic region is the receiving part of the synapse comprising thousands of proteins forming an elaborate and dynamically changing network indispensable for the molecular mechanisms behind fundamental phenomena such as learning and memory. Despite the growing amount of information about individual protein-protein interactions in this network, these data are mostly scattered in the literature or are stored in generic databases that are not designed to display aspects which are fundamental to understanding postsynaptic functions. To overcome these limitations we collected postsynaptic protein-protein interactions (PPIs) together with a high amount of detailed structural and biological information and launched a freely available resource, the Postsynaptic Interaction Database (PSINDB) to make these data and annotations accessible. PSINDB includes tens of thousands of binding regions together with structural features mediating and regulating the formation of PPIs, annotated with detailed experimental information about each interaction. PSINDB is expected to be useful for numerous aspects of molecular neurobiology research, from experiment design to network and systems biology-based modeling and analysis of changes in the protein network upon various stimuli. PSINDB is available at <http://psindb.itk.ppke.hu/>.

Introduction

Synapses are communication points between neurons responsible for transducing information. Synapses can be broadly categorized into presynaptic sites releasing neurotransmitters, and postsynaptic sites taking in chemical signals. Cellular and organism level events, such as neuronal development, memory and long-term potentiation, strengthen synaptic connections [1]. Excitatory synapses contain a special morphological unit called postsynaptic density (PSD), that localizes directly under the membrane of the receiving cell [2]. The PSD is composed of thousands of highly conserved proteins; however the exact distribution and organization of components change during neuronal development leading to morphological differences in the PSDs of different brain regions [3]. Mutations in proteins of the PSD are responsible for severe neurological and psychiatric diseases [4], many of them likely to be heritable as suggested by a growing number of evidence, such as in the case of Autism Spectrum Disorder [5,6].

Protein-protein interactions (PPIs) play an essential role both in the maintenance of synaptic plasticity and in disease emergence. PSD proteins exhibit a high degree of multivalency presenting multiple binding sites stacked into single proteins. This multivalency aids the formation of an elaborate and dynamic network via highly modular architecture, provided by distinct structural and functional elements [7]. Intrinsically disordered regions (IDRs) were shown to be critical components for network assembly in the PSD [8], often by containing short linear motifs mediating transient interactions [9], post-translational modification sites regulating interactions as switches [10] and more. These regions play crucial roles in forming higher-order assemblies of the PSD, such as fuzzy complexes [11] and membraneless organelles via liquid-liquid phase separation [12].

Although there is a huge amount of information available about PPIs formed in the postsynapse, they are either scattered in the literature or they are collected in databases that were not prepared to store extended information that helps us to better understand the nature of synapse formation and function at the molecular level. Here we present the Postsynaptic Interaction Database (PSINDB, <https://psindb.itk.ppke.hu/>) aiming to provide a fuller and more complex picture of the postsynaptic protein network. PSINDB is a comprehensive database focused on postsynaptic proteins, containing a high number of collected, manually curated or derived interactions with precisely defined binding regions, enriched by various structural features of the constituting proteins. We believe that PSINDB will help scientists to get closer to the understanding of the complex nature of synaptic function.

Methods

Selection of postsynaptic proteins

We downloaded the human, mouse and rat reference proteomes (2021_April release) from UniProt [13]. We assigned orthologous proteins using the OMA database [14] and gene names. To define postsynaptic proteins the following sources were utilized: SynptomeDB [15], Genes2Cognition [16], SynGo [17] and GeneOntology postsynapse/postsynaptic density terms (and their child terms) [18].

Manual curation

For the first step in the curation we extensively searched PubMed and Google scholar for relevant publications. Since many of these publications were already included in other PPI databases, we only used articles that were not processed elsewhere. The curation work was done adhering to the HUPO-PSI Molecular Interaction (MI) workgroup standards, and at least MIMix curation level [19] was used, while terms describing interactions were taken from the PSI-MI ontology (<https://www.ebi.ac.uk/ols/ontologies/mi>). We considered every described interaction between rat, mouse and human proteins, then mapped them back to the human reference proteome.

Data integration

We integrated binary PPI information from the IntAct [20], BioGrid [21] and STRING [22] databases. Furthermore, we derived interaction data from PDB [23]: first we reconstructed the full structure using the BIOMT lines, then Voronota [24] was used to calculate intermolecular surfaces and to identify interacting residues.

We have predicted or downloaded the following protein structure-related information: transmembrane segments (Human Transmembrane Proteome database [25], liquid-liquid phase separation (PhaSePro [26]), short linear motifs (Eukaryotic Linear Motifs database, [27]), phosphorylation (UniProt annotations), domains (PFAM [28]), coiled-coils (Deepcoil [29] and UniProt annotations) disordered regions and disordered binding regions (IUPred2A [30]) and disease causing germline mutations classified in Disease Ontology [31]. Protein alignments between isoforms were done using ClustalOmega [32].

Results

The homepage of PSINDB database is available at the URL: <https://psindb.itk.ppke.hu/>. We have created an interactive graphical user interface for the visualization of interaction data.

Two types of searches can be performed on the site I) the protein search will result in a list of proteins II) the interaction search will result in a list of interacting partners. In both cases the gene name, entrez and HUGO gene ID, UniProt ID and various aliases can be used as a query.

Clicking results from the protein search opens the individual protein pages, which includes a graphical scheme of the corresponding protein along with ten sections summarizing structural, interaction and network data (Figure 1). The sections are the as follows: 1) Evidence for PSD localization (with link to the source database) 2) Function (short description mirrored from UniProt); 3) Protein features (structural and interaction information, featuring interacting regions collecting binding regions of all partners); 4) Binary interactions with known binding regions (interacting regions with each partner one by one); 5) Network (a matrix-like representation of partners, where the depth of the color correlates with the strength of experimental evidence for the interaction); 6) Isoforms (alignment of isoforms highlighting segments not included in splice variants making it easily comparable with the location of the binding regions) 7) Disease-causing germline mutations (mutations and diseases associated to the protein, together with the partners where binding sites overlap with the mutated residue) 8) Linear motifs (together with the partners where binding sites overlap with the mutated residue) 9) Fingerprint (percentages of GeneOntology molecular function, GeneOntology biological process and disease ontology terms shared within the network of binary interaction partners of the protein) 10) All partners (including those not having postsynaptic annotations, yet the interaction indicate the partner may have the same localization).

The interaction page can be accessed from the Protein page by clicking on the partner, or from the interaction search page. The interaction page contains the definition and details about the experiment (Figure 1), including, but not limited to the type of interaction, interaction detection method, host organism, experimental and biological role of participants and more. Binding regions for each experiment are also displayed where available.

In addition, we prepared several predefined subsets of postsynaptic proteins/interactions - including scaffolding proteins, proteins involved in phase separation, hub proteins and more - that may be interesting to a broad range of users.

PSINDB core data is currently available to download in the PSI-MI mitab format [33], with further download options planned in the near future.

Figure 1: Layout of the PSINDB database with A4_human protein as an example. Protein page 1) Evidences for PSD localization 2) Function 3) Protein features 4) Binary partners with known interacting regions 5) Network 6) Isoforms 7) Disease causing germline mutations 8) Short linear motifs 9) Fingerprint 10) All partners. Interaction page 11) experimental details.

Conclusion

PSINDB is a comprehensive database of postsynaptic protein interactions, featuring detailed experimental information, binding regions of partners, and structural and biological annotations. The current version of PSINDB stores 16,618 unique binary interactions between 2,210 proteins, with 484,314 experimental evidence from 22,892 research articles. We expect that PSINDB will be a valuable resource for researchers investigating the molecular mechanisms of synaptic signal transduction, function, regulation and misregulation. In particular, our detailed data on protein-protein interaction sites can substantially help to understand how different complexes from the same components can be assembled and how these can interconvert into one another. PSINDB can serve as a resource for network and systems biology-based models of postsynaptic organization and function. PSINDB is available at <https://psindb.itk.ppke.hu/>

Acknowledgement

Support from the National Research, Development and Innovation Office through grant OTKA 124363 to Z.G. is acknowledged. BM has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 842490 (MIMIC).

References

- [1] T.M. Boeckers, The postsynaptic density, *Cell Tissue Res.* 326 (2006) 409–422.
- [2] M. Sheng, E. Kim, The postsynaptic organization of synapses, *Cold Spring Harb. Perspect. Biol.* 3 (2011). <https://doi.org/10.1101/cshperspect.a005678>.
- [3] S. Vucetic, H. Xie, L.M. Iakoucheva, C.J. Oldfield, A.K. Dunker, Z. Obradovic, V.N. Uversky, Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions, *J. Proteome Res.* 6 (2007) 1899–1916.
- [4] T. Kaizuka, T. Takumi, Postsynaptic density proteins and their involvement in neurodevelopmental disorders, *J. Biochem.* 163 (2018) 447–455.
- [5] L. de la Torre-Ubieta, H. Won, J.L. Stein, D.H. Geschwind, Advancing the understanding of autism disease mechanisms through genetics, *Nat. Med.* 22 (2016) 345–361.
- [6] X. Liu, T. Takumi, Genomic and genetic aspects of autism spectrum disorder, *Biochem. Biophys. Res. Commun.* 452 (2014) 244–253.
- [7] A. Kiss-Tóth, L. Dobson, B. Péterfia, A.F. Ángyán, B. Ligeti, G. Lukács, Z. Gáspári, Occurrence of Ordered and Disordered Structural Elements in Postsynaptic Proteins Supports Optimization for Interaction Diversity, *Entropy* . 21 (2019). <https://doi.org/10.3390/e21080761>.
- [8] M.S. Cortese, V.N. Uversky, A.K. Dunker, Intrinsic disorder in scaffold proteins: getting more from less, *Prog. Biophys. Mol. Biol.* 98 (2008) 85–106.
- [9] T. Mittag, R. Parker, Multiple Modes of Protein-Protein Interactions Promote RNP Granule Assembly, *J. Mol. Biol.* 430 (2018) 4636–4649.
- [10] M.D. Ehlers, Activity level controls postsynaptic composition and signaling via the ubiquitin-proteasome system, *Nat. Neurosci.* 6 (2003) 231–242.
- [11] E. Kwok, D.J. Rodriguez, J. Kremerskothen, A. Nyarko, Intrinsic disorder and amino acid specificity modulate binding of the WW2 domain in kidney and brain protein (KIBRA) to synaptopodin, *J. Biol. Chem.* 294 (2019) 17383–17394.
- [12] Z. Feng, X. Chen, M. Zeng, M. Zhang, Phase separation as a mechanism for assembling dynamic postsynaptic density signalling complexes, *Curr. Opin. Neurobiol.* 57 (2019) 1–8.

- [13] UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, *Nucleic Acids Res.* 49 (2021) D480–D489.
- [14] M. Zahn-Zabal, C. Dessimoz, N.M. Glover, Identifying orthologs with OMA: A primer, *F1000Res.* 9 (2020) 27.
- [15] M. Pirooznia, T. Wang, D. Avramopoulos, D. Valle, G. Thomas, R.L. Haganir, F.S. Goes, J.B. Potash, P.P. Zandi, SynaptomeDB: an ontology-based knowledgebase for synaptic genes, *Bioinformatics.* 28 (2012) 897–899.
- [16] M.D.R. Croning, M.C. Marshall, P. McLaren, J.D. Armstrong, S.G.N. Grant, G2Cdb: the Genes to Cognition database, *Nucleic Acids Res.* 37 (2009) D846–51.
- [17] F. Koopmans, P. van Nierop, M. Andres-Alonso, A. Byrnes, T. Cijssouw, M.P. Coba, L.N. Cornelisse, R.J. Farrell, H.L. Goldschmidt, D.P. Howrigan, N.K. Hussain, C. Imig, A.P.H. de Jong, H. Jung, M. Kohansalnadehi, B. Kramarz, N. Lipstein, R.C. Lovering, H. MacGillavry, V. Mariano, H. Mi, M. Ninov, D. Osumi-Sutherland, R. Pielot, K.-H. Smalla, H. Tang, K. Tashman, R.F.G. Toonen, C. Verpelli, R. Reig-Viader, K. Watanabe, J. van Weering, T. Achsel, G. Ashrafi, N. Asi, T.C. Brown, P. De Camilli, M. Feuermann, R.E. Foulger, P. Gaudet, A. Joglekar, A. Kanellopoulos, R. Malenka, R.A. Nicoll, C. Pulido, J. de Juan-Sanz, M. Sheng, T.C. Südhof, H.U. Tilgner, C. Bagni, À. Bayés, T. Biederer, N. Brose, J.J.E. Chua, D.C. Dieterich, E.D. Gundelfinger, C. Hoogenraad, R.L. Haganir, R. Jahn, P.S. Kaeser, E. Kim, M.R. Kreutz, P.S. McPherson, B.M. Neale, V. O’Connor, D. Posthuma, T.A. Ryan, C. Sala, G. Feng, S.E. Hyman, P.D. Thomas, A.B. Smit, M. Verhage, SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse, *Neuron.* 103 (2019) 217–234.e4.
- [18] The Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong, *Nucleic Acids Res.* 47 (2019) D330–D338.
- [19] S. Orchard, L. Salwinski, S. Kerrien, L. Montecchi-Palazzi, M. Oesterheld, V. Stümpflen, A. Ceol, A. Chatr-aryamontri, J. Armstrong, P. Woollard, J.J. Salama, S. Moore, J. Wojcik, G.D. Bader, M. Vidal, M.E. Cusick, M. Gerstein, A.-C. Gavin, G. Superti-Furga, J. Greenblatt, J. Bader, P. Uetz, M. Tyers, P. Legrain, S. Fields, N. Mulder, M. Gilson, M. Niepmann, L. Burgoon, J. De Las Rivas, C. Prieto, V.M. Perreau, C. Hogue, H.-W. Mewes, R. Apweiler, I. Xenarios, D. Eisenberg, G. Cesareni, H. Hermjakob, The minimum information required for reporting a molecular interaction experiment (MIMIx), *Nat. Biotechnol.* 25 (2007) 894–898.
- [20] S. Kerrien, B. Aranda, L. Breuza, A. Bridge, F. Broackes-Carter, C. Chen, M. Duesbury, M. Dumousseau, M. Feuermann, U. Hinz, C. Jandrasits, R.C. Jimenez, J. Khadake, U. Mahadevan, P. Masson, I. Pedruzzi, E. Pfeifferberger, P. Porras, A. Raghunath, B. Roechert, S. Orchard, H. Hermjakob, The IntAct molecular interaction database in 2012, *Nucleic Acids Res.* 40 (2012) D841–6.
- [21] R. Oughtred, C. Stark, B.-J. Breitkreutz, J. Rust, L. Boucher, C. Chang, N. Kolas, L. O’Donnell, G. Leung, R. McAdam, F. Zhang, S. Dolma, A. Willems, J. Coulombe-Huntington, A. Chatr-Aryamontri, K. Dolinski, M. Tyers, The BioGRID interaction database: 2019 update, *Nucleic Acids Res.* 47 (2019) D529–D541.
- [22] D. Szklarczyk, A.L. Gable, K.C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N.T. Doncheva, M. Legeay, T. Fang, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets, *Nucleic Acids Res.* 49 (2021) D605–D612.
- [23] S.K. Burley, C. Bhikadiya, C. Bi, S. Bittrich, L. Chen, G.V. Crichlow, C.H. Christie, K. Dalenberg, L. Di Costanzo, J.M. Duarte, S. Dutta, Z. Feng, S. Ganesan, D.S. Goodsell, S. Ghosh, R.K. Green, V. Guranović, D. Guzenko, B.P. Hudson, C.L. Lawson, Y. Liang, R. Lowe, H. Namkoong, E. Peisach, I. Persikova, C. Randle, A. Rose, Y. Rose, A. Sali, J. Segura, M. Sekharan, C. Shao, Y.-P. Tao, M. Voigt, J.D. Westbrook, J.Y. Young, C. Zardecki, M. Zhuravleva, RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences, *Nucleic Acids Res.* 49 (2021) D437–D451.
- [24] K. Olechnovič, C. Venclovas, Voronota: A fast and reliable tool for computing the

- vertices of the Voronoi diagram of atomic balls, *J. Comput. Chem.* 35 (2014) 672–681.
- [25] L. Dobson, I. Reményi, G.E. Tusnády, The human transmembrane proteome, *Biol. Direct.* 10 (2015) 31.
- [26] B. Mészáros, G. Erdős, B. Szabó, É. Schád, Á. Tantos, R. Abukhairan, T. Horváth, N. Murvai, O.P. Kovács, M. Kovács, S.C.E. Tosatto, P. Tompa, Z. Dosztányi, R. Pancsa, PhaSePro: the database of proteins driving liquid-liquid phase separation, *Nucleic Acids Res.* 48 (2020) D360–D367.
- [27] M. Kumar, M. Gouw, S. Michael, H. Sámano-Sánchez, R. Pancsa, J. Glavina, A. Diakogianni, J.A. Valverde, D. Bukirova, J. Čalyševa, N. Palopoli, N.E. Davey, L.B. Chemes, T.J. Gibson, ELM-the eukaryotic linear motif resource in 2020, *Nucleic Acids Res.* 48 (2020) D296–D306.
- [28] S. El-Gebali, J. Mistry, A. Bateman, S.R. Eddy, A. Luciani, S.C. Potter, M. Qureshi, L.J. Richardson, G.A. Salazar, A. Smart, E.L.L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S.C.E. Tosatto, R.D. Finn, The Pfam protein families database in 2019, *Nucleic Acids Res.* 47 (2019) D427–D432.
- [29] J. Ludwiczak, A. Winski, K. Szczepaniak, V. Alva, S. Dunin-Horkawicz, DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences, *Bioinformatics.* 35 (2019) 2790–2795.
- [30] B. Mészáros, G. Erdos, Z. Dosztányi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding, *Nucleic Acids Res.* 46 (2018) W329–W337.
- [31] L.M. Schriml, C. Arze, S. Nadendla, Y.-W.W. Chang, M. Mazaitis, V. Felix, G. Feng, W.A. Kibbe, Disease Ontology: a backbone for disease semantic integration, *Nucleic Acids Res.* 40 (2012) D940–6.
- [32] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Molecular Systems Biology.* 7 (2011) 539. <https://doi.org/10.1038/msb.2011.75>.
- [33] L. Perfetto, M.L. Acencio, G. Bradley, G. Cesareni, N. Del Toro, D. Fazekas, H. Hermjakob, T. Korcsmaros, M. Kuiper, A. Lægreid, P. Lo Surdo, R.C. Lovering, S. Orchard, P. Porras, P.D. Thomas, V. Touré, J. Zabolos, L. Licata, CausalTAB: the PSI-MITAB 2.8 updated format for signalling data representation and dissemination, *Bioinformatics.* 35 (2019) 3779–3785.