Final report

The goal of the project was the exploration of the structure and function of the intrinsically disordered parts of the postsynaptic scaffold protein GKAP. Then postsynaptic density is an elaborate supramolecular machine with a high number of multivalent interactions, and the almost fully disordered GKAP protein is one of the most abundant components. To date, there are no detailed structural models for full-length GKAP and its unstructured segments, which make up the majority of the protein, have not been explored in detail. Especially, the possible relationships between the free and bound states of multiple binding sites is yet to be explored. We have focused on four regions, the dynein light chain-binding region, the GK-.binding region, the C-terminal GH1 domain, and the N-terminal variable part. It should be stressed that most of these constructs are novel in the sense that no reported previous protocol is available, and thus, each of them needs specific optimization. Due to the pandemic situation, we had limited access to our laboratory for extended periods, and thus resolving the difficulties with the samples took much more time than could be anticipated when planning the project, even with the generously provided two extension periods (7 months total). Nevertheless, we made every effort to proceed with sample preparation and characterization and achieved important milestones, detailed below, that pave the way to complete the yet missing investigations in the foreseeable future, suitable for description in several publications. We have also performed a number of computational studies and have also implemented novel methods for the characterization of GKAP, its participation in the formation of protein complexes and the organization of the postsynaptic density in general. We have also investigated proteins and segments that are binding partners of GKAP.

Experimental characterization of GKAP segments and their interacting partners

The dynein light chain-binding region of GKAP harbors two binding sites for the dynein light chain DYNLL8, itself a dimer. Our main goal was to characterize the binding interaction and the emerging complex. To this end, we have expressed the corresponding segment of GKAP (from residue 691 to residue 747 in rat GKAP, UniProt ID P97836) in unlabeled, ^{15}N-labeled and ^{15}N, ^{13}C-labeled form.

Our collaborating partner, Prof. Perttu Permi (University of Jyväskylä, Finland) has successfully obtained 2D and 3D NMR spectra for resonance assignment. Our assignment of the first construct indicated a largely disordered structure, consistent with the prior predictions. We have also measured the complex formed by GKAP and DYNLL8 using both ^{15}N-labeled GKAP segment and unlabeled DYNLL8 as well as ^{15}N-labeled DYNLL8 and unlabeled GKAP. Thus, we were able to do a preliminary analysis of the binding sites on both partners. Our initial results strongly suggest that both predicted binding sites on GKAP are functional in the construct (Figure 1).

However, analysis of CB resonances also indicated that both cysteine residues in the GKAP construct are in the oxidized form, forming either an intra- or an intermolecular disulfide bridge. As this most likely does not reflect the biologically relevant state of the intracellular protein GKAP, we set out to adjust the buffer conditions with the addition of a reducing agent. Based on

our own and previously obtained experience with DYNLL8, our choice for this was TCEP. Experiments to confirm the relevance of our results described above under conditions more relevant to the physiological one are under way. Unfortunately, several samples proved to be unstable despite the success with the first constructs and conditions. We expect qualitatively similar results as those described above but confirmation with the reduced form is necessary before reliable structural models can be built and published. The experiments with ^{15}N-labeled DYNLL and unlabeled GKAP also confirmed the binding with the involvement of the previously described partner binding region in DYNLL8. Currently, our latest sample is under investigation at our partner's laboratory in Jyväskylä.

The GK-binding region of GKAP bears five segments that can presumably interact with the GK domain of the PSD-95 protein. We have expressed various constructs covering this region, and from these, GBR1-3, containing the first three binding sites (numbering from the N-terminus) turned out to be the most promising one. However, even for this construct, obtaining samples of quantity and purity suitable for NMR measurements turned out to be a challenge. In collaboration with Prof. Permi's laboratory, we were recently able to perform initial measurements confirming the disordered nature of the region (Figure 1). However, GK binding only occurs when the corresponding binding regions on GKAP are phosphorylated. We have successfully optimized conditions for phosphorylation with the kinase CAMKII.

We have also expressed the GK domain of PSD-95 in ^{15}N and ^{13}C-labeled forms and the spectra obtained confirm that the construct is folded. The signal quality of the 3D spectra recorded so far for resonance assignment did not allow a more detailed characterization of this construct yet as we were directing our efforts to the GBR1-3 segment.

The C-terminal GH1 domain of GKAP is the only one for which a previously published structure is available, albeit one stabilized *via* a fusion protein. The C-terminus of GKAP interacts with the PDZ domain of proteins in the Shank family.

We have expressed the Shank1 PDZ domain in ^{15}N and ^{15}N, ^{13}C-labeled forms and have performed resonance assignment on spectra recorded at the University of Debrecen. We did not aim at full structure determination as a number of X-ray structures are available for this domain, rather focused on information that can specifically be obtained *via* solution-state NMR spectroscopy. To our knowledge, this is the first NMR investigation of this domain. We could obtain an almost complete backbone resonance assignment and could also identify more than 90% of the side-chain CB chemical shifts. Analysis of sequence-neighborhood corrected secondary chemical shifts confirmed that our structure is folded. It also indicated the presence of a short helical segment within the β2-β3 loop, which, to our knowledge, has only been observed in one of the most recently determined X-ray structures, and our results confirmed its presence in solution. In addition, the rotational diffusional correlation time calculated from heteronuclear ^{1}H-^{15}N relaxation data is consistent with the presence of the monomeric form of the domain in solution. This is an important finding as previously multiple dimerization modes were suggested, primarily on the basis of crystal structures. Our results have been published in an open access paper and the chemical shift and heteronuclear relaxation data have been deposited in BioMagResBank under accession no. 51126.

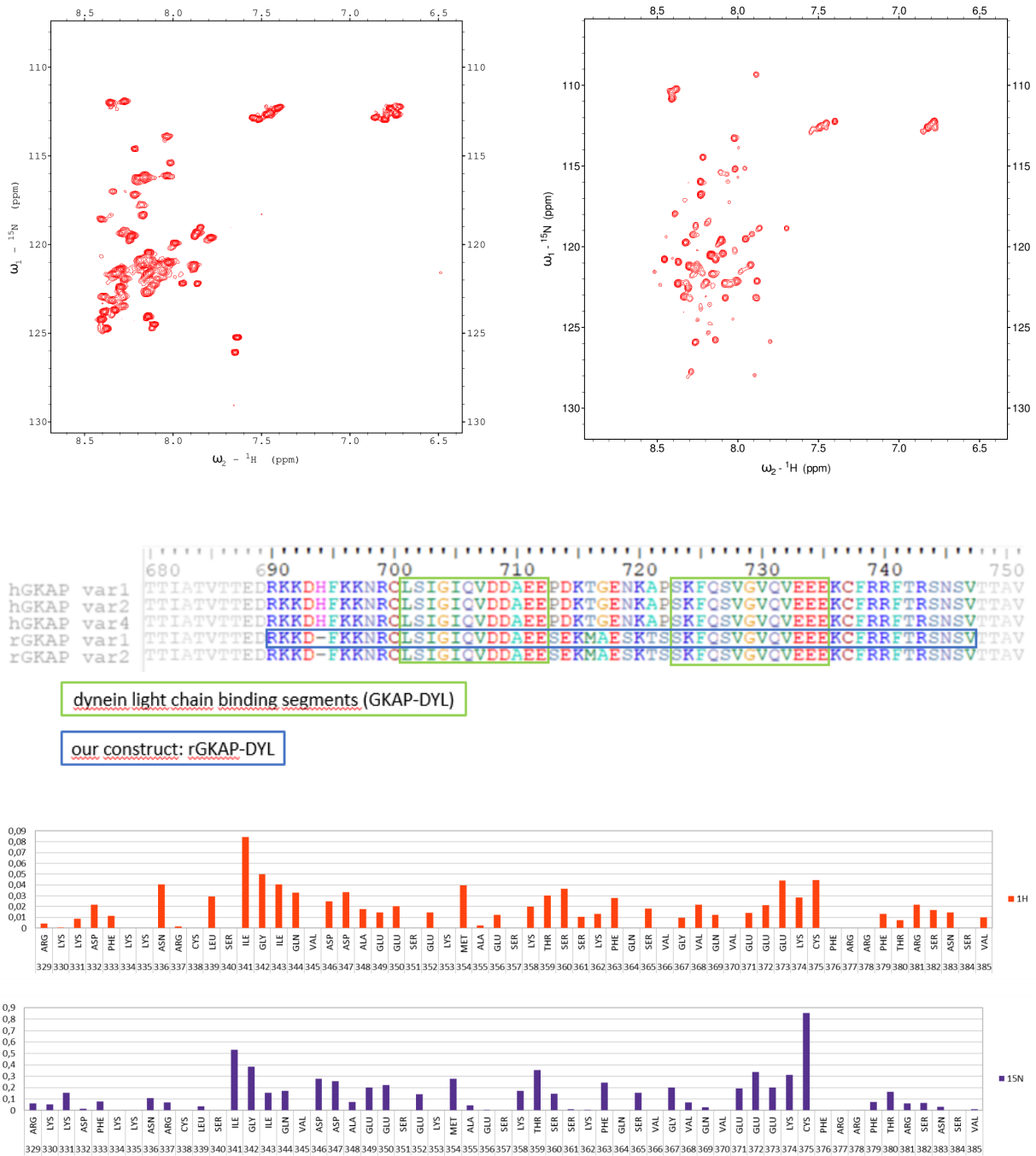

Figure 1. Top left: ^{1}H-^{15}N HSQC spectrum of the DYNLL8-binding region of GKAP. Top left: ^{1}H-^{15}N HSQC spectrum of the GK-binding region (top right) of GKAP. Middle: sequence alignment of the dynein-binding region of GKAP. Bottom: ^{1}H and ^{15}N chemical shift perturbations observed upon the addition of sub-stoichiometric amount of DYNLL8.

With pull-down assays, biolayer interferometry and ITC measurements we have also confirmed the interaction between Shank1 PDZ and the C-terminal segment of GKAP using constructs of various lengths. Preliminary NMR measurements also provided proof of the interaction, but the observed spectral changes were not straightforward to interpret. To assess the structural aspects of the binding interaction, additional experiments with multiple labeling schemes are needed. Preparation of a selected construct for detailed NMR investigation of the interaction is currently underway.
In addition to the above, we have also initiated a systematic analysis of the N-terminal segment of GKAP that is specific in the longest isoform and its function is elusive. We have prepared a construct library and could select a few of them that might be suitable for further analysis.

Corresponding publications, including selected conference presentations:

Anna Sánta, András Czajlik, Gyula Batta, Bálint Péterfa, Zoltán Gáspári:
Resonance assignment of the Shank1 PDZ domain
BIOMOLECULAR NMR ASSIGNMENTS (2022), DOI: 10.1007/s12104-022-10069-4

Anna Sánta, Melinda Keresztes, Eszter Nagy-Kanta, József Hegedüs, Viktor Farkas, Gyula Batta, Bálint Péterfia, Zoltán Gáspári: Characterization of the interaction between Shank-PDZ and GKAP-GH1 domains (poster), FEBS Congress, 2021; FEBS Open Bio 11: S1

Eszter Nagy-Kanta, Anna Sánta, Fanni Farkas, Perttu Permi, Maarit Hellmann, Helena Tossavainen, Zoltán Gáspári, Bálint Péterfia: NMR-based structural characterization of the post-synaptic density scaffold protein GKAP (poster), FEBS Congress, 2021; FEBS Open Bio 11: S1

Fanni Farkas, András Czajlik, Gyula Batta, Bálint Péterfia, Zoltán Gáspári: Atomic-level investigations of the PSD-95 GK domain and its interaction with GKAP (poster), FEBS Congress, 2021; FEBS Open Bio 11: S1

Development of computational tools for the generation and analysis of disordered protein segments of GKAP

Intrinsically disordered segments like those in GKAP can not be described by structural models consisting of only a single conformer. Therefore, ensemble-based models are considered more suitable descriptions. In such models, no single conformer is expected to account for the experimental data, but the full ensemble as a whole corresponds to the measured parameters. Although there are multiple methods published in the literature, our experience showed that both the availability and flexibility of these tools can be limited, considering the scope of our analysis. Our intention was to develop a simple, open-source, and flexible pipeline for the generation of structural pools of intrinsically disordered proteins. The generated structural ensembles can then be subjected to the desired selection process to obtain an ensemble

reflecting the experimentally measured parameters. The developed DIPEND pipeline is implemented in Python and uses open-source third-party software like ChimeraX and GROMACS. It also utilizes the widely used Scrwl4 program and the publicly available neighbor-dependent Ramachandran database developed in the laboratory of Prof. Dunbrack. The idea behind using this database is that the generated structural pool already reflects structural preferences characteristic of the sequence. To make the tool more flexible, a user-added biasing distribution can be added in a residue-specific manner to direct the sampling of the conformational space toward conformations matching the experiments better. This makes it possible to use relatively small structure pools of 5-10,000 conformers that are both large enough for selection and at the same time samples the most relevant conformations of the molecule investigated. To facilitate the generation of longer structures, an “unknotting” algorithm has been added that only minimally perturbs the distribution of the backbone dihedrals yet helps to generate clash-free segments that are encountered with increasing probabilities for segments above 60-70 residues and/or with high proline content. The DIPEND pipeline is freely available on GitHub at: https://github.com/PPKE-Bioinf/DIPEND. DIPEND is described in an open-access publication.

To obtain sub-ensembles that reflect the actually measured experimental parameters, we have used our previously developed CoNSEnsX^{+} server which is capable of performing a deterministic selection using a user-defined set of available experimental parameters with different relative weights. We have updated the server with the addition of the analysis of neighbor-dependent secondary chemical shifts that is essential for the analysis of intrinsically unstructured protein segments where the chemical shifts are expected to be close to their random coil values and small deviations bear important information on local structural preferences. In addition, we have implemented support for the externally available Bayesian Maximum Entropy (BME) tool, developed in the group of Prof. Lindorff-Larsen, that can reweight structural ensembles to optimize their correspondence to experimental data. The latest version of CoNSEnsX^{+} is available as a public webserver at consensx.itk.ppke.hu and the source code can be found on GitHub at: https://github.com/PPKE-Bioinf/consensx.itk.ppke.hu

The usage and novel developments of CoNSensX+ were described in a book chapter in the series of “Methods in Molecular Biology” as well as in the DIPEND paper.

The combination of DIPEND and CoNSEnsX^{+} is currently used to build structural models of the GKAP segments investigated, which will be used for building full-length models of the GKAP protein (Figure 2.).

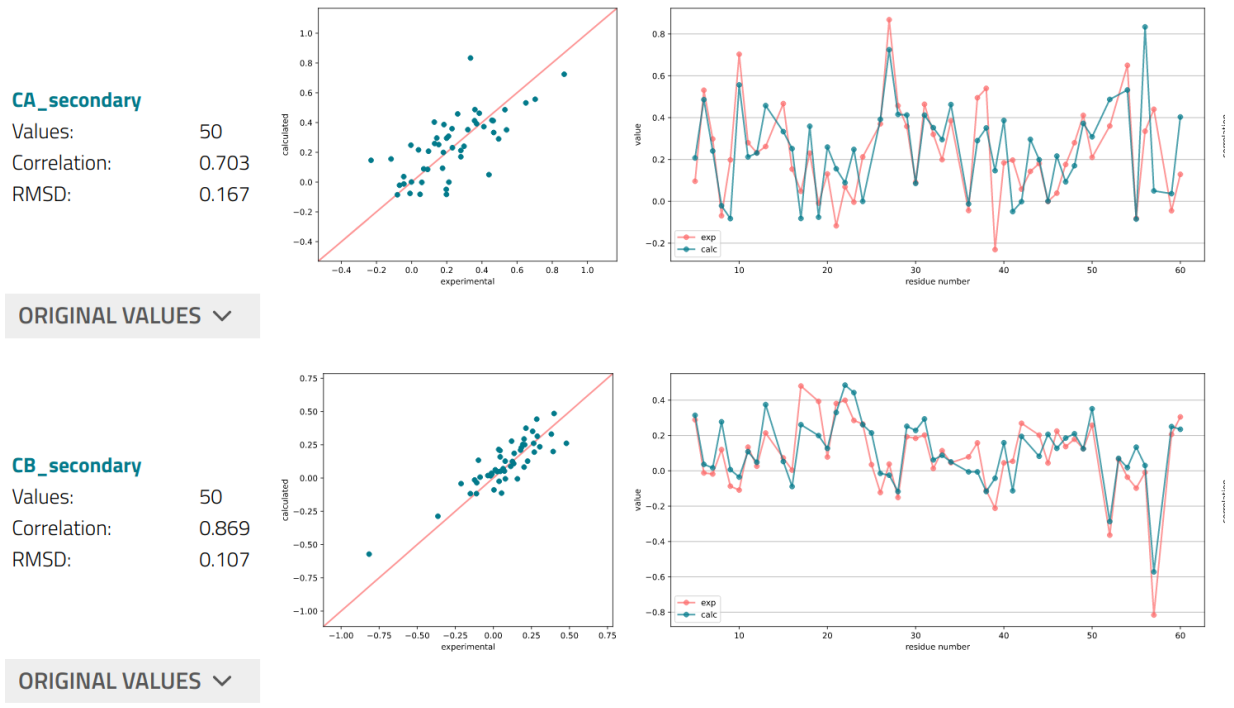

Figure 2. Snapshot of a CoNSEnsX⁺ -based analysis of secondary Cα and Cβ chemical shifts in selected models from an exploratory ensemble of the GKAP DYNLL8-binding segment generated with DIPEND. Correlation and RMSD values indicate acceptable correspondence between the ensemble and the assigned chemical shifts, that can form the basis of further optimizations

Corresponding publications:

Zita Harmat, Dániel Dudola, Zoltán Gáspári:
DIPEND: An Open-Source Pipeline to Generate Ensembles of Disordered Segments Using Neighbor-Dependent Backbone Preferences
BIOMOLECULES 11 : 10 p. 1505 (2021)

Dániel Dudola, Bertalan Kovács, Zoltán Gáspári:
Evaluation and Selection of Dynamic Protein Structural Ensembles with CoNSEnsX±
In: Gáspári, Zoltán (ed.) Structural Bioinformatics : Methods and Protocols, pp. 241-254. Springer (Boston), 2020

Computational analysis of the PDZ domains of PSD-95

PSD-95 is both the most abundant postsynaptic protein and one of the main interaction partners of GKAP. To understand this interaction in full atomistic detail, it is desirable to build full-length structural models of both molecules. To be able to build such models with sufficient accuracy, the dynamic nature of the domains and segments should be explicitly modeled. To our knowledge, there are no detailed NMR data available for the GK domain, and our own measurements indicated the need for further sample optimization which could not yet be done. However, the PDZ1-PDZ2 domain pair and, especially, the PDZ3 structure of PSD-95 is subject to intensive research and a number of NMR parameters and observations are available. Based on these, we have performed a detailed computational analysis of the PDZ1-2 domain pair as well as the PDZ3 domain. For all three domains the free and ligand-bound states were modeled and S^2 backbone order parameters were incorporated in multi-replica simulations with an in-house modified version of GROMACS. For the PDZ3 domain, full-length and truncated variants, missing the C-terminal α-helix, not commonly found in PDZ domains, were compared to get insight into the previously suggested allosteric regulation *via* this particular segment. In addition, we have generated small ensembles of further PDZ domains where chemical shift data were available. In such cases, the ensemble was selected with CoNSEnsX[+] from a pool generated by unrestrained molecular dynamics simulations. Our results suggest that the loops flanking the peptide binding site have a key role in mediating the interactions between the PDZ1 and PDZ2 domains. As motions of these loops are dependent upon the presence or absence of ligands, we could establish a clear structure-based explanation of the coupled nature of the intra- and interdomain motions in this PDZ tandem. Furthermore, we suggested a dynamic model of intradomain allostery of PDZ3 based on population shifts along the two most important motions identified by principal component analysis. A comparative analysis of multiple PDZ domains pinpointed a region outside the immediate peptide-binding cleft that is responsible for most of the structural variation between PDZ domains. In addition, we made a quite unexpected observation, namely, that the opening-closing motions of loops around the binding site are not uniformly associated with the free and ligand-bound forms but whether the open or the closed state is present in the complex form is domain-dependent.
Our results were described in two open-access papers.

Corresponding publications:
Dániel Dudola, Anett Hinsenkamp, Zoltán Gáspári:
Ensemble-Based Analysis of the Dynamic Allostery in the PSD-95 PDZ3 Domain in Relation to the General Variability of PDZ Structures
INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES 21 : 21 Paper: 8348 (2020)

Bertalan Kovács, Nóra Zajácz-Epresi, Zoltán Gáspári:
Ligand-dependent intra- and inter-domain motions in the PDZ12 tandem regulate binding interfaces in postsynaptic density protein-95
FEBS LETTERS 594 : 5 pp. 887-902. , 16 p. (2020)

Large-scale computational analysis of postsynaptic proteins and their complexes

GKAP is a major postsynaptic scaffold protein with multiple binding sites for at least two of its known partner molecules. This multivalent nature of the protein greatly increases the number of possible complexes formed with the participation of GKAP. Inspired by this feature, we performed computational investigations to assess whether this kind of multivalency is a general feature of postsynaptic proteins. Our results indicate that postsynaptic proteins have characteristic structural features that discriminate them from other protein sets, although many postsynaptic proteins are also involved in other biological processes unrelated to synaptic signal transduction. Our findings suggest that indeed many PSD proteins bear characteristics that promote interaction diversity.

Our survey of postsynaptic proteins and their interactions revealed that general protein:protein interaction databases either lack data on a number of postsynaptic proteins and/or do not provide detailed enough annotation to represent potentially multivalent interactions. Currently available databases of (post)synaptic proteins, on the other hand, do not contain interaction data. Therefore, we decided to build a protein:protein interaction database specialized for postsynaptic proteins (Figure 3). To this end, we compiled postsynaptic proteins from existing databases and annotations, interactions described in the IntAct database and complemented these data with a thorough survey of the relevant literature. Through this, we could add several hundred interactions not annotated in any database we are aware of, and, wherever available, we also provide information on the binding regions at different levels, in accordance with state-of-the art standards outlined for PPI databases. The developed database, PSINDB, is freely available for online browsing and download at: https://psindb.itk.ppke.hu/

To assess the role of multivalency in postsynaptic protein complex formation, we have also performed an exploratory analysis of complex formation of seven major postsynaptic proteins including GKAP. We have used the protein complex simulation approach Cytocast and interaction details, as well as protein abundance data sets available in the literature. Our main premise was that according to the synaptomic theory, the identity of the synapses is a key determinant of their functional role, and, in turn, the latter is determined largely by the protein interactions and the distribution of resulting complexes. rather than the protein abundances themselves. Our simulations reveal that there is a non-trivial relationship between the abundance of the individual proteins and the distributions of the emerging protein complexes. In other words, there are cases where a relatively small alteration in the protein abundances causes considerable redistribution of the protein complexes. We have set up a simple web interface that can be used to reproduce our calculations, available at http://psdcomplexsim.cytocast.com/.

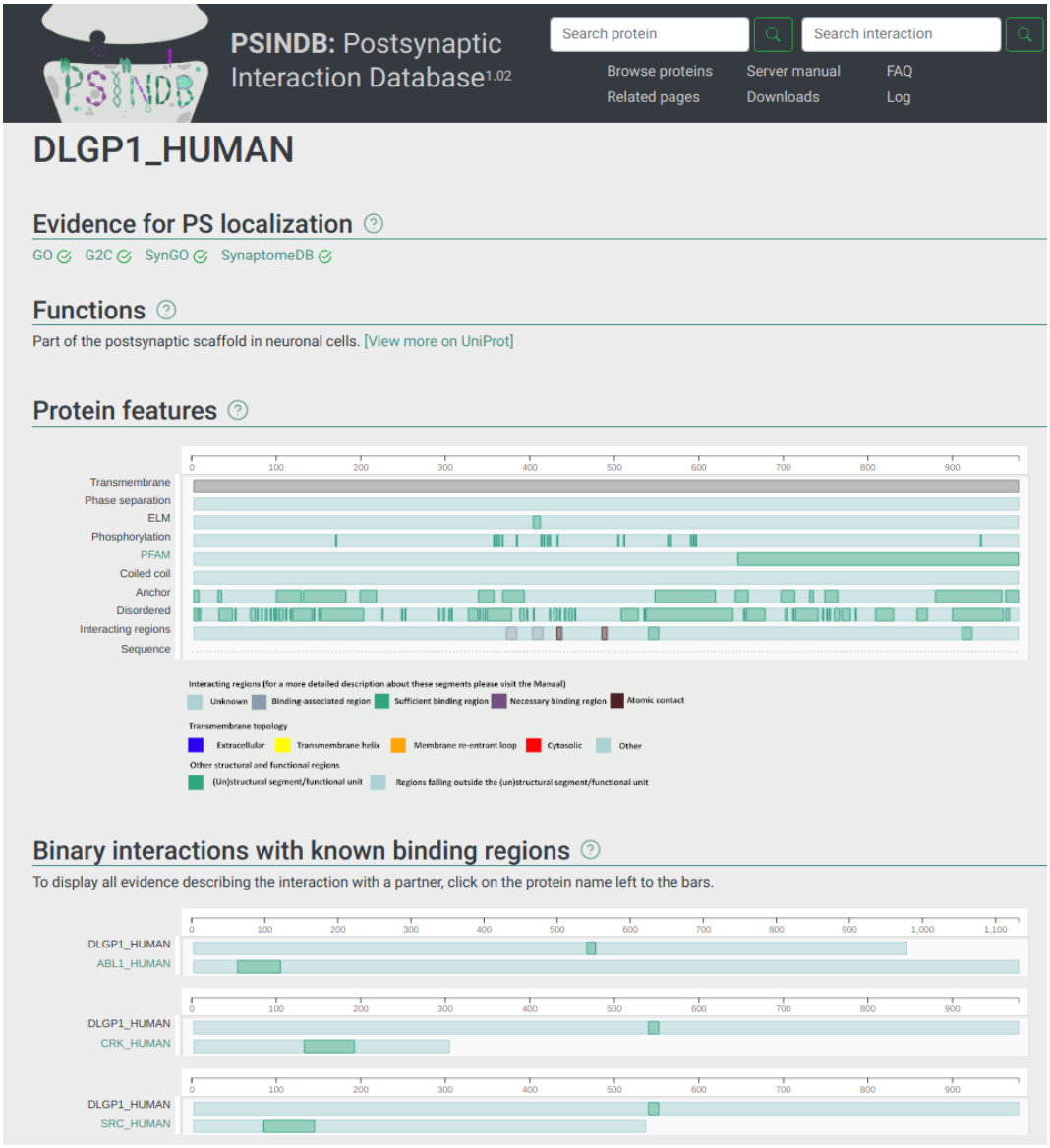

Figure 3. A screenshot showing part of the information on GKAP in the PSINDB database

Corresponding publications:

Annamária Kiss-Tóth, László Dobson, Bálint Péterfia, Annamária F. Ángyán, Balázs Ligeti, Gergely Lukács, Zoltán Gáspári:
Occurrence of Ordered and Disordered Structural Elements in Postsynaptic Proteins Supports Optimization for Interaction Diversity
ENTROPY 21 : 8 Paper: 761 , 16 p. (2019)

Zsófia Kálmán, Dániel Dudola, Bálint Mészáros,, Zoltán Gáspári, László Dobson:
PSINDB: The postsynaptic protein-protein interaction database
DATABASE-JOURNAL OF BIOLOGICAL DATABASES AND CURATION 2022 : 2022 Paper: baac007 (2022)

Marcell Miski, Bence Márk Keömley-Horváth, Dorina Rákóczi Megyeriné, Attila Csikász-Nagy, Zoltán Gáspári:
Diversity of synaptic protein complexes as a function of the abundance of their constituent proteins: A modeling approach
PLOS COMPUTATIONAL BIOLOGY 18 : 1 Paper: e1009758 (2022)

Additional computational analyses with relevance to postsynaptic proteins

Initially, we have set out to investigate the properties of postsynaptic proteins and some of their special structural elements. However, some of these studies led to the recognition of more general phenomena but still with relevance to PSD proteins.

Our first survey of disease-causing mutations in postsynaptic proteins quickly redirected our focus to coiled coils where a unique pattern of such mutations became apparent. A more detailed analysis led to the recognition of the enrichment of disease-containing mutations in the N-terminal region of coiled coils. A possible explanation for this can lie in the cooperative folding of these structures for which disruption of interactions near the N-terminus can be more detrimental than a downstream perturbation.

Liquid-liquid phase separation (LLPS) is a phenomenon best described for RNA-binding and nuclear proteins but also for postsynaptic assemblies. Our previous results on single alpha helices (SAHs) indicated the association of this structural element with RNA binding. Therefore, we have analyzed the possible association of the presence of SAHs and other segments with high density of charged residues with LLPS. We found a strong association between the presence of SAHs and phase separation propensity of proteins. However, a closer look indicates that this is a result of a negative association, namely, that proteins not prone to LLPS are depleted in SAHs and other charge-dense elements. It is important to stress that although in some cases the charge-dense region coincides with the segment responsible for triggering LLPS, our interpretation of the observed association is that regions with high charge density and/or regularly alternating charge patterns (like SAHs) can effectively function both in the solution and the condensed phase by maintaining strong ionic interactions within and between protein molecules. Interestingly, postsynaptic proteins are not particularly enriched in such regions and therefore their participation in LLPS is likely dominated by other types of interactions and sequence features.

The recognition of specific subclasses of so-called low complexity regions is crucial for the proper characterization of protein sequences and their functional relevance. Low-complexity sequences with a repeating motif can form fibrillar structures like collagen helices, coiled coils or even SAHs, whereas other segments with similar amino acid distributions but lacking regularity are often intrinsically disordered. Thus, using the proper nomenclature to describe sequence

and structure-based features and their interdependence is of high importance in protein analysis and annotations. We have participated in a large international team effort aiming at the clarification of the relevant terminologies and description of the relevant features with examples.

Corresponding publications:

Zsófia E. Kálmán, Bálint Mészáros, Zoltán Gáspári, László Dobson:
Distribution of disease-causing germline mutations in coiled-coils implies an important role of their N-terminal region
SCIENTIFIC REPORTS 10 : 1 Paper: 17333 (2020)

András L. Szabó, Rita Pancsa, Anna Sánta, Zoltán Gáspári:
Charged sequence motifs increase the propensity towards liquid–liquid phase separation
FEBS LETTERS 598 : 8 pp. 1013-1028. 16 p. (2022)

Pablo Mier, [...] Zoltán Gáspári, [...] et al.:
Disentangling the complexity of low complexity proteins
BRIEFINGS IN BIOINFORMATICS 21 : 2 pp. 458-472. , 15 p. (2020)