

RESEARCH

Open Access



Simulated complexes formed from a set of postsynaptic proteins suggest a localised effect of a hypomorphic Shank mutation

Marcell Miski¹, Áron Weber^{1,2}, Krisztina Fekete-Molnár¹, Bence Márk Keömley-Horváth^{1,2}, Attila Csikász-Nagy^{1,2*} and Zoltán Gáspári^{1*}

Abstract

Background The postsynaptic density is an elaborate protein network beneath the postsynaptic membrane involved in the molecular processes underlying learning and memory. The postsynaptic density is built up from the same major proteins but its exact composition and organization differs between synapses. Mutations perturbing protein: protein interactions generally occurring in this network might lead to effects specific for cell types or processes, the understanding of which can be especially challenging.

Results In this work we use systems biology-based modeling of protein complex distributions in a simplified set of major postsynaptic proteins to investigate the effect of a hypomorphic Shank mutation perturbing a single well-defined interaction. We use data sets with widely variable abundances of the constituent proteins. Our results suggest that the effect of the mutation is heavily dependent on the overall availability of all the protein components of the whole network and no trivial correspondence between the expression level of the directly affected proteins and overall complex distribution can be observed.

Conclusions Our results stress the importance of context-dependent interpretation of mutations. Even the weakening of a generally occurring protein: protein interaction might have well-defined effects, and these can not easily be predicted based only on the abundance of the proteins directly affected. Our results provide insight on how cell-specific effects can be exerted by a mutation perturbing a generally occurring interaction even when the wider interaction network is largely similar.

Keywords Protein complex, Mutation, Systems biology, Protein interaction network, Protein:protein interaction, Gillespie algorithm, Postsynaptic density, Binding affinity

*Correspondence:

Attila Csikász-Nagy
csikasznagy@gmail.com
Zoltán Gáspári
gaspari.zoltan@itk.ppke.hu

¹Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary

²Cytocast Hungary Kft, Budapest, Hungary



© The Author(s) 2024, corrected publication 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

The synaptic theory states that the identity of the synapses in different cell types is a key component in establishing the complexity of brain functions, including learning and memory [1, 2]. There are a number of experimental observations indicating that the postsynaptic density (PSD), utilizing the same major constituent proteins, can be highly variable depending on the abundance of its individual protein components [3]. The PSD is also capable of dynamic reorganization during the circadian cycle and upon stimuli [4], and its flexibility to do so has been proposed to play a key role in synaptic plasticity and network rewiring [5]. Despite our knowledge in a number of pairwise protein: protein interactions between postsynaptic proteins [6], the specific large-scale organization of the PSD is still largely elusive.

Mutations identified in proteins of the PSD in various neurological conditions might directly affect specific protein: protein interactions. Although the same proteins and interactions can occur in virtually all postsynaptic compartments, the effect of mutations is often specific to a brain region, to specific interaction partners or leads to the impairment or gain of defined functions instead of leading to the dysfunction of neural transmission in all or most cells [7]. This behavior is expected to be only interpretable by taking into account the complexity of the *in vivo* conditions.

The Shank protein family contains large modular scaffold proteins with both globular and long intrinsically disordered regions [8]. These proteins establish a number of diverse interactions with various postsynaptic proteins. Mutations affecting the availability and/or the structure of Shank proteins have been linked to many conditions [9] from autism spectrum disorder (ASD) [10] to Phelan-McDermid Syndrome (PMS) [11]. In these conditions, Shank3 haploinsufficiency [12], caused by either the complete loss of a copy of the gene or by the presence of a function-affecting mutation, is the most prevalent cause, although mutations in the Shank1 and Shank2 genes can also cause similar, although typically milder, phenotypes.

Besides state-of-the-art microscopic techniques [13, 14], simulation approaches might contribute to our more detailed understanding of the supramolecular structure of the PSD and its changes upon stimuli and mutations. We have previously described extensive simulations on protein complex formation using a simplified model of the PSD containing 7 major proteins. These simulations revealed that the correspondence between protein component abundance and the distribution of the complexes formed is nontrivial [15]. Specifically, the classification of the simulated PSDs based on the abundance of constituent proteins can largely differ from the classification based on the resulting protein complex distributions. Importantly, protein complexes represent a feature

supposedly more closely linked to the biological function of the network than simple protein abundance. Our data set included instances of the same or similar brain regions with different expression levels of the individual proteins, and as such our previous work also models the effect of abundance-affecting mutations.

In this work, we address the question of whether hypomorphic mutations, i.e. those reducing but not completely abolishing a function, specifically a binding interaction, can cause measurable effects in our PSD model system. We have chosen the Shank1 PDZ domain as a model system because of its well-studied nature [16]. Mutations in this domain have been linked to ASD such as R736Q [10] and also have been identified in various cancers [17]. As a model hypomorphic mutation, we have chosen one that can be estimated to cause the decrease of a specific binding interaction between 2- and 10-fold. We use the same model system as in our previous work, composed of seven major PSD proteins, except for replacing Shank3 with Shank1 (NMDAR, AMPAR, PSD-95, SynGAP, GKAP, Homer1, Shank1) [15]. As previously, we performed our simulations on more than 500 brain regions with different protein levels [18].

For all simulations, the same set of interactions is defined, except for the one affected by the mutation, giving rise to the „wild type” and „mutated” scenarios. For each region, the input is the abundance of each protein and the output is the number of the different protein complexes formed. Our results suggest that weakening a single well-defined interaction does not affect the overall distribution of complexes in most investigated brain regions. However, in a small number of cases, the most informative protein complexes - defined based on their contribution to the overall diversity of the PSD complexes - show a significant change in their abundance. These results indicate that even when the same set of proteins is involved, the biological effect of a mutation can be highly specific depending on the cellular context.

Methods

Overview of the simulations

The aim of our simulations was to estimate the distributions of protein complexes formed by a set of proteins with different abundances (copy numbers). The set of proteins was a highly simplified PSD model including 7 major proteins, two receptors (AMPA, NMDAR) and five scaffold proteins (Homer1, GKAP, SynGAP, and Shank1) with well-defined interactions between them (Tables S1, S2, S3).

For the simulation of protein complex formation, we have used an agent-based simulation tool (Cytocast), also applied in our previous study [15]. The underlying principle is the Gillespie algorithm, a related implementation of which was used to predict COVID-19 outcomes [19].

A previous implementation (SiComPre) was shown to effectively model protein complex distribution in whole cells [20].

Data sets, setup and reproducibility

Overall, we have simulated protein complexes in 524 different data sets originating from 27 brain region types, as in our previous work [15] (Table S4). The source data sets contain mRNA expression levels [18] that we have translated into protein abundance by linear scaling of the values to the copy number of PSD-95 proteins as described previously [15] (Table S5). We note that mRNA levels can be used to approximate cellular protein abundance [21, 22]. Still, ours is a highly simplified approach that does not take any other factors like protein degradation into account, but is sufficient for our modeling purposes to obtain data sets with sufficient variation. The average copy numbers of the individual proteins, serving as input for the simulations, show a high variance (Table 1). It should be noted that no information about the individuals from which the data are derived is available, whether they can be considered as in a healthy or diseased condition. The raw input abundances can be found in the Supplementary material (Table S5). Dissociation constants were taken from the data described in the literature [23]. In our simulations, the dissociation constant was implemented by adjusting the unbinding rate as a result of a mutation, while not changing the binding rates. To estimate the reproducibility of the simulations, 40 simulations were performed for each data set. This number was deemed sufficient in terms of reproducibility, as with 40 repetitions we have reached a variance of 1 that did not improve upon more repetitions. In addition, Shapiro-Wilk test for normality, performed for the abundance of the most informative complex (see below) with Benjamini-Hochberg correction [24] indicated that the normal distribution of complex abundances can be assumed (can not be rejected) in all 524 regions when using 40 simulations.

To make testing our calculations and setups possible, we have set up a web interface, available at <https://psd-complexsim.cytocast.com>.

Table 1 Statistics of input protein abundance derived from mRNA expression levels from reference [18]

Protein	Minimum	Maximum	Average	Deviation
NMDAR	0	95	16.84	17.87
AMPA	1	688	126.81	78.39
PSD-95	36	1067	328.09	159.21
SynGAP	11	359	1102.21	60.13
GKAP	2	367	82.62	65.50
Shank1	1	388	69.20	59.40
Homer1	1	124	21.98	17.20

Choice and modeling of the Shank1 R743H mutation

In order to use a well-studied interaction, we have chosen the Shank1 PDZ domain and its interaction with the C-terminus of GKAP. The Shank1 PDZ is a globular domain that has been characterized extensively [25]. Moreover, several mutations in the Shank1 PDZ domain has been observed in patients with Autism Spectrum Disorder (ASD) [10]. There are also mutations for this domain listed in the COSMIC database [17], and as these are available in an organized format, we decided to use this source to explore several mutations.

Our goal was to model a moderate yet measurable effect of a hypomorphic mutation that weakens but does not abolish binding. It is not trivial to estimate the perturbation effect of a mutation, especially for one outside the primary ligand binding site. Therefore, we have used the change in domain stability as predicted by the NeEMO method [26] and have used this change to estimate the weakening of the interaction by assuming less stable apo and holo structures.

For our modeling purposes, we have chosen the R743H mutation for which we have estimated a 5.5-fold decrease in the binding affinity, modeled as a 5.5-fold increase for the dissociation rate of the Shank1 PDZ: GKAP interaction in our setup. This value is regarded as a good compromise between a minimally observable 2-fold change and changing the value by one order of magnitude, a more pronounced change. The arginine affected by this mutation is located on the C-terminal end of helix 2, whereas the R736Q mutation, described in ASD, alters an arginine at the N-terminus of the same helix (Fig. 1a). Both arginines point away from the immediate ligand binding site, thus, both mutations are expected to perturb the binding interaction indirectly.

Analysis of the simulation results

In our system composed of multivalent proteins, the identity of the protein complexes formed is not only determined by their composition but also the exact topology, i.e. how the constituent proteins interact with each other, rather also by the binding sites actually participate in the interactions. Complexes with the same protein composition but different binding patterns should therefore be distinguished during the analysis of the simulation results.

All protein complexes observed were identified and enumerated, and each was assigned a unique identifier. The simulated results for each input data set can be represented by a point in a multidimensional space where each coordinate represents the abundance of a given protein complex. Thus, a protein complex distribution can be described as the linear combination of the emerging protein complexes (Eq. 1):

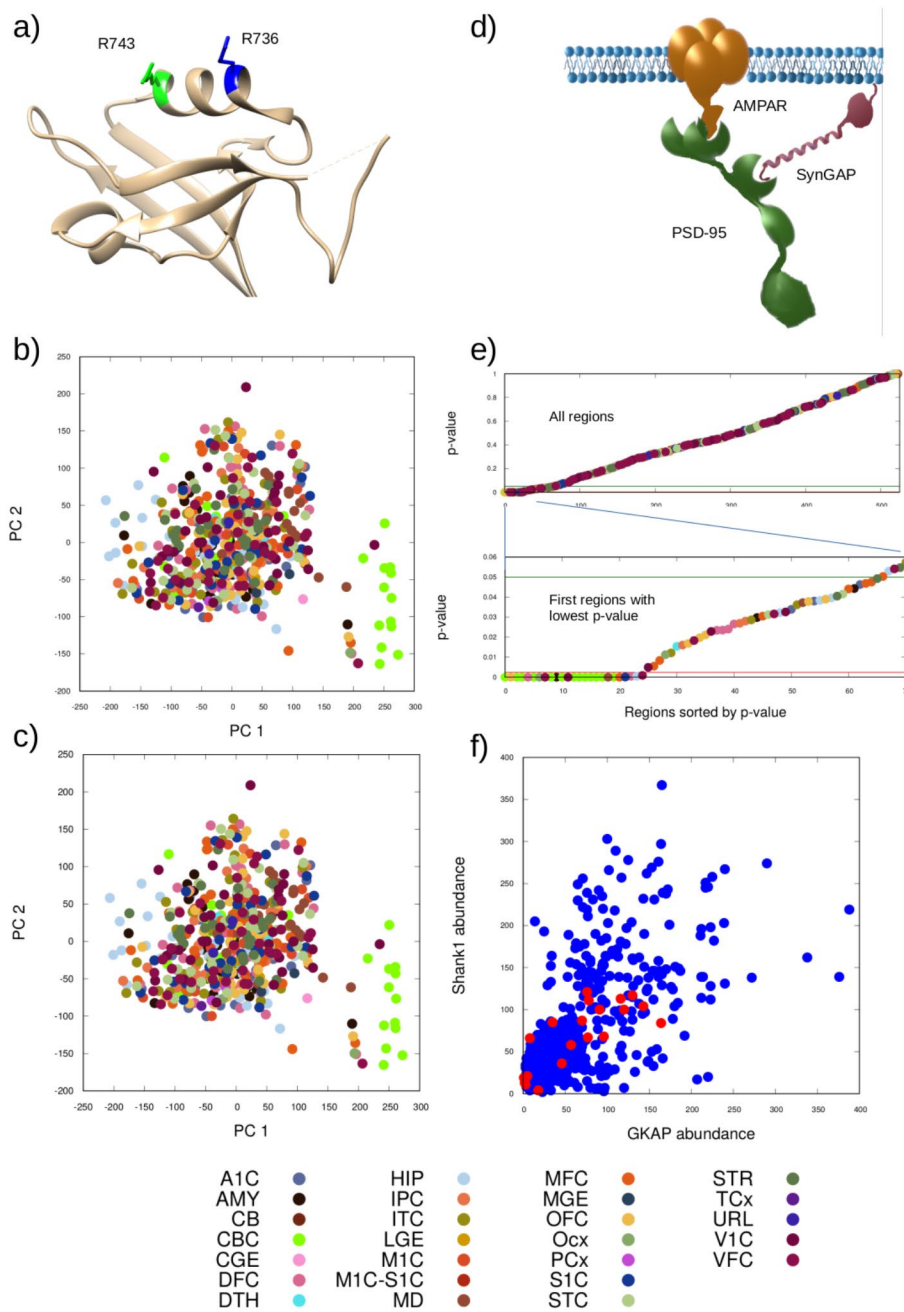


Fig. 1 (a) Position of the mutation selected (R743H, green) and a similar one reported in ASD (R736Q, blue) on the ribbon representation of the Shank1 PZ domain (PDB ID 6YWZ). Both arginines are located on the $\alpha 2$ helix flanking the ligand binding groove. Principal component analysis of the obtained protein complex distributions for (b) the wild-type and (c) the mutant scenarios investigated. Different colors denote different brain regions according to the key at the bottom. (d) Schematic depiction of the most informative complex according to the PCA (AMPA/PSD-95/SynGAP). (e) P-values describing the change upon the mutation relative to the wild-type, the value for the most informative complex is shown in increasing order from left to right, colored by the region type (key at the bottom) The green line denotes the 0.05 significance limit while the red line the limit of 0.0024 obtained using the Benjamini-Hochberg correction. (f) Abundance of Shank1 and GKAP, the two proteins in the interaction affected by the mutation, in the input data sets. Red circles indicate data sets where the abundance of the most informative complex changed significantly in the output using the Benjamini-Hochberg correction

$$c \in R^n, c = \sum_{i=1}^n \alpha_i p_i \quad (1)$$

Where a_i is the abundance of complex i and p_i is a base (unit) vector along axis i representing the copy number of the i th protein complex.

Each region can exhibit a different complex distribution. To analyze the differences between regions and simulations, we used principal component analysis. In the PCA outputs, the points represent the different complex abundances for each region in the dimension-reduced space allowing the largest differences between the brain regions to be visualized. Originally, each region could be assigned a vector with as many dimensions as complexes, with a value for the abundance of each complex. Our data matrix X used for PCA in our study has dimensions of 524 (regions) by 222,784 (complex types), with a rank of 524. This setup ensures that the PCA can appropriately capture the variability and structure within our data, with 524 independent components reflecting the rank of the matrix. To get more insight into the role of individual complexes in the PCA, we have determined the importance of each complex by combining two measures obtained from the standard principal component analysis: (1) the fraction of variability explained by the given principal component and (2) the contribution of the abundance of the protein complex investigated to each of the PCs. In this way it is possible to calculate how relevant and informative the original base vectors (representing complex abundances) are in the comparison of brain regions.

$$r \in R^n, r = \sum_{i=1}^n \frac{\lambda_i}{\sum_{j=1}^n |u_{ij}|} u_i \quad (2)$$

Where r is an n -dimensional vector containing the relevance of (variance explained by) each complex, λ_i is the relevance of the i -th eigenvector, u_{ij} is the j -th coordinate (component in the original space with a dimension for each complex) of the i -th eigenvector u_i .

Statistical comparison of complexes and regions

A pairwise T-test is a classical statistic usually chosen when only one change - here the mutation - is created in the system and the question is how the change affects the mean value. The null hypothesis is that the mean abundance of a complex (in the wild-type and mutant scenario) is the same within a certain level of significance. Thus the alternative hypothesis is that the mean abundance of a given complex formed from the protein set with the wild-type and the mutant proteins differ significantly.

The calculations were performed as implemented in the scipy package (method `scipy.stats.ttestrel`), where the t-score is calculated as in Eq. 3 [27]:

$$t_{score} = \frac{mean(a - b)}{\sigma} \quad (3)$$

where a and b are 40-length arrays of the abundances of the given complex for wild-type and mutant simulations respectively and σ is the standard error.

The p-value is calculated from the t-score based on the alternative hypothesis type that is two-sided:

$$p = 2cdf_{t,39}(-|t_{score}|) \quad (4)$$

where cdf is the cumulative density function for the T-test with the degree of freedom 39 (for 40 repetitive simulations).

Given that the abundance of each complex is averaged separately during the simulation, the T-test shows which hypothesis we can accept for the abundance only of the given complex (p_i).

Results

We ran several simulations based on the Gillespie algorithm. The inputs of the simulations are protein abundances and parameters for simulating protein complex formation dynamics, along with the descriptions of the possible pairwise interactions. Protein abundance data, being highly variable as shown in Table 1 along with at PCA analysis in Figure S1. The outputs are protein complexes and their abundances for each brain region. In our simulation outputs, we have observed the formation of 222,784 different complexes.

Identification of the most informative complex

Principal component analyses (PCA, Fig. 1b, c) show which complexes are the most informative from the point of view of distinguishing brain regions. The first two principal components for both the wild-type and mutant scenarios cover 44% and 24% of the full variance of the outputs, respectively. The first principal axis is dominated by the abundance of the AMPAR/PSD-95 (id:12) complex, whereas the second one by the PSD-95/SYNGAP (id:8) complex.

The most informative complex overall - with the highest contribution considering all principal components and the variance explained by them - is AMPAR/PSD-95/SYNGAP (id:5) (Fig. 1d) with a contribution of 19%. In comparison, the importance of the average complex is very small, approximately $4.48e-06$ and the median is $9.98e-08$. This is because of the high number of possible complexes that can form.

Complex composition and relevance

Common proteins in the complexes formed are PSD-95, AMPAR and SYNGAP. PSD-95 appears in all the top 10 most informative complexes, indicating its central role in complex formation. AMPAR is present in 6 out of the 10 complexes, showing its frequent participation, whereas SYNGAP is involved in 5 of the complexes, highlighting its significance. This pattern suggests a common scaffold or core structure. Despite the recurrence of certain proteins, the specific combinations and additional proteins like NMDAR and GKAP create diverse structures. Notably absent from the top 10 complexes are Shank1 and Homer1, proteins known for their ability to polymerize or heterotetramerize. Their absence suggests that Shank1 and Homer1 and the complexes that have Shank1 and Homer1 are involved in forming much larger complexes. These larger complexes tend to have lower individual occurrences, leading to reduced overall variance and relevance in this specific analysis.

Overall complex distribution is primarily determined by protein availability

Principal component analysis of the simulation results for the wild-type and mutant scenarios show a very similar overall picture. The two PCA plots can directly be compared as the axes are the same even in the two independent PCA outputs. The data points corresponding to the wild-type and mutant cases move only minimally relative to each other (Table S6). The average distance between wild-type and mutant regions is 1.5 ± 0.8 .

We note that our PCA results do not generally separate the source brain regions with the exception of the cerebellar cortex-type regions that are concentrated at a distinct region from the others (Fig. 1b, c).

Our results suggest that the overall protein complex distribution is determined by the availability of the individual proteins and the presence of a weakening mutation does not cause substantial global effects. This is in line with the system retaining its general functionality. In order to analyze the effect of the mutation in more detail, we have investigated the abundance of the individual protein complexes.

The significance of changes in the formation of the most informative complex does not seem to be closely related to the abundance of Shank1 and GKAP, the interaction of which is directly affected by the mutation (Fig. 1f). The only observable pattern is that no significant changes are found in regions with high abundance of both proteins. On the other hand, many regions with low or moderate abundance of these two proteins also do not show a significant change. This observation reinforces our conclusion that the availability of partners in the wider interaction network has a profound effect on complex formation.

Weakening the Shank: GKAP interaction causes subtle effects in well-defined regions

To analyze the effect of the mutation introduced, we have compared the complex distributions in each region with the wild-type simulations. At the level of individual complexes, there are some exhibiting significant changes in the mutated scenario relative to the wild-type one.

For larger supercomplexes, even small numerical change in their abundance can be significant due to the very low probability of their formation. Therefore, in the case of supercomplexes, even the appearance of one complex can be considered relevant.

Importantly, the p-values of the most informative complex AMPAR/PSD-95/SynGAP drop below 0.0024 for some of the resulting data sets (Fig. 1e, Table S8). This is somewhat surprising as this complex does not contain either of the partners of the interaction affected by the mutation. Thus, it is worth investigating how the abundance of the individual partner proteins Shank1 and GKAP influence the changes observed in the formation of AMPAR/PSD-95/SynGAP complex. We have plotted the input abundance of Shank1 and GKAP vs. the p-value of the most informative complex (Fig. 1e). It is apparent that significant changes are not confined to either the high- or the low-abundance regions of the two affected proteins (Fig. 1f), suggesting that the complex interplay between the interactions of multiple proteins is behind the observed phenomenon.

As a result of the mutation, we would expect the complexes in the layer above GKAP, i.e. those containing the membrane receptors and PSD-95, to become more favored since the interaction between GKAP and Shank1 connects these complexes to the larger supercomplexes where Shank1 polymerizes. This effect is observed only in the cerebellar cortex regions and to a small extent. Some example abundances are shown in Figure S2.

The abundances change from zero to 1 for the complexes SynGAP/PSD-95/GKAP (id:9) and AMPAR/PSD-95/GKAP (id:15). However, overall complex abundances remain highly similar to the wild-type even in the regions with the smallest p-values.

Curiously, the regions with the lowest p-values all belong to the cerebellar cortex. The structure and levels of the cerebellum are significantly different compared to the cerebrum, which also means differences in the main neuron types [28]. The differences have already been demonstrated for different Shank3, Shank2 abundances in different layers in the cerebellum [29], and the different contributions of the cerebellum to ASD [30]. Naturally, we do not claim that the observed low p-values directly reflect these aspects, but it cannot escape our attention.

To get further insight into the changes of complex abundance, we have selected two regions: H376. IIB.53M1C-S1C exhibits the lowest nonzero p-value for

the most informative complex (AMPA/PSD-95/SynGAP) (Figure S3).

In these two regions, the average abundance of the complex AMPAR/PSD-95/SynGAP (id:5) changes from 306 to 301 upon the mutation. In addition, the abundance of the related complex AMPAR/PSD-95/SynGAP/GKAP increases. Thus, the decrease can be partially attributed to the fact that the complex AMPAR/PSD-95/SynGAP (id:5) associates with GKAP with a higher probability than in the case of the wild-type. The greater availability of uncomplexed GKAP can be explained by the weakened Shank1:GKAP connection.

On the other hand, for the two regions with the highest p-values (H376.XI.50HIP and H376.VIII.53MD), the abundance of the complex AMPAR/PSD-95/SynGAP does not change at all (Figure S4).

Discussion

Justification of our approach

Our model of only seven PSD proteins and without any specific spatial organization is definitely a highly simplified one that is far from the actual biological complexity of the postsynapse. In addition, for simplicity, we consider a situation that corresponds to a homozygous scenario, i.e. where either only wild-type or mutant Shank1 is present but not both. Last but not least, we have modeled only one well-defined effect of the mutation, ignoring possible pleiotropic effects like the alteration of the expression level of multiple proteins as observed for several Shank mutations [31]. Thus, it is not expected that the obtained protein complex distributions can be directly compared to the *in vivo* situations. Modeling all these aspects with acceptable accuracy would require much more data than currently available. However, we argue that our model system, focusing on a well-defined set of major PSD proteins and interactions is complex enough to capture general aspects of the behavior of elaborate protein networks with a multitude of binding interactions while remaining manageable in terms of data analysis as the number of possible protein complexes is not extremely high. On the other hand, mechanistic linking of genotypes with phenotypes, with different genotypes leading to similar phenotypes, is only possible via a combination of experimental data and modeling approaches.

The fact that the distribution of complex formation is not distinguishable in different brain regions and does not significantly change upon mutation does not necessarily indicate that the results are random. Here are some key points to consider: The cerebellum being isolated, we found distinguishable patterns, suggesting that the results are not entirely random. If the results were purely random, we would expect more significant changes across all regions, including the cerebellum. The physical

binding model used in the study is deterministic, meaning that the processes it simulates are governed by specific physical laws. Only the diffusion process involves random elements, which are based on probabilistic associations and dissociations. If the dependence on inputs were truly random, it would undermine the reproducibility of the results. However, the model's consistency with physical laws supports its reliability. The reproducibility of the results across different simulations and conditions strengthens the argument against randomness. Consistent findings across various simulations suggest that the observed patterns are driven by underlying biological processes rather than random fluctuations.

Experimental investigation of the dependence of protein complex formation on the availability and binding properties of its constituent proteins would be a challenge. *In vitro* reconstruction of multicomponent systems is far from routine but not unprecedented [32]. However, quantitative analysis of a number of different species formed would require elaborate methods, most likely based on mass spectrometry. In addition, the use of full-length proteins, especially membrane receptors might not be feasible, requiring the design of constructs containing only the relevant interaction sites. In an *in vivo* setting, immunoprecipitation or high-resolution microscopy like 3D z-stack STORM could be used at the level of individual synapses [33]. However, for a multicomponent complex the different fluorescent labeling of the constituent proteins might be an issue along with the quantification of the individual proteins and conducting the experiments in diverse brain regions.

Weakening a specific interaction can cause limited but significant changes

Mechanistic linking mutations to the phenotypes they cause is many times a non-trivial task, especially when the mutation perturbs a highly complex protein network. This phenomenon is well known in the case of neurodevelopmental disorders, where similar observed phenotypes can be caused by a number of different mutations. For example, a recent recommendation for Phelan-McDermid syndrome puts emphasis on the underlying genetic cause as the phenotypes are largely non-specific and can generally occur in a number of neurodevelopmental diseases [11].

The specific effect of mutations can be enigmatic, especially when they affect proteins present in many different tissues. This is especially true for cell types in which even the major partners and interactions are expected to be the same. The diversity of neurons in terms of the different abundance of postsynaptic proteins offers a unique opportunity to explore the effect of specific mutations in a complex but still simplified multicomponent system having the same set of building blocks.

Our simulation-based approach, focusing on the formation of protein complexes as defined by the abundance of their constituent proteins and their interactions, reveals that the effect of a mutation weakening a specific interaction heavily depends on the availability of all interaction partners in the system. The complex interdependence of the interactions leads to a scenario where the overall changes in protein complex distributions are generally subtle, the formation of only a few complexes are significantly affected and this effect can be confined to a well-defined set of cells with specific protein abundance. While common wisdom could suggest that protein complexes containing the mutated proteins are most affected and in cells where these are abundant, our results indicate that cells with the lowest number of affected proteins can also be among the vulnerable ones, and the protein associates mostly affected are linked only indirectly to the actually weakened interaction. Our simulations suggest that the cerebellum may be an involved brain region (Figure S4), aligning with findings in the literature [28–30].

Conclusions

Our results suggest that even for a ubiquitous occurring interaction, effects can be highly dependent on the wider context, including the availability of all components in a larger protein-protein interaction network. Our modeling, however simplified, indicate that the affected cells can not be easily predicted based purely on the abundance of the interaction partners. Modeling different scenarios with distinct protein abundances also suggest that mutations that (also) affect protein expression levels may have a more serious effect than those that moderately perturb a given interaction. In addition, the effect of redundancy in the system (e.g. Shank1-Shank3) might not only cause individual isoforms directly take over each other's roles, but rather they reduce intermediary effects, e.g. here, the change in the frequency of the most important complex would always be moderated by the presence of a protein with a binding pattern similar to that of the mutant. We believe that with the availability of more data from single-cell studies and combining these with simulations like the one presented here, we might get closer to the understanding of cell-to-cell variability in healthy and diseased conditions.

Abbreviations

PSD	Postsynaptic density
ASD	Autism spectrum disorder
PMS	Phelan-McDermid Syndrome
PCA	Principal component analysis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12868-024-00880-1>.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

Not applicable.

Author contributions

ZG and AC-N conceived and supervised the study, MM and ZG designed simulations, MM, ÁW, KF-M and BMK-H performed data collection, setup, running and analysis of simulations, MM, GZ and AC-N wrote the manuscript.

Funding

Open access funding provided by Pázmány Péter Catholic University. The authors acknowledge the support of the National Research, Development and Innovation Office – NKFIH – through grants no. TKP2021-EGA-42 and OTKA 137947.

Open access funding provided by Pázmány Péter Catholic University.

Data availability

All data used for the simulations during this study are included in this published article and its supplementary information files. To make testing our calculations and setups possible, we have set up a web interface, available at <https://psdcomplexsim.cytocast.com>. Other data are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

ÁW, BK-H and AC-N reporting that they have financial and business interests in Cytocast Kft. that may be affected by the research reported in the enclosed paper. All other authors declared no competing interests.

Received: 9 February 2024 / Accepted: 28 June 2024

Published online: 06 July 2024

References

1. Grant SGN. The synaptomic theory of behavior and brain disease. *Cold Spring Harbor Symp Quant Biol.* 2018a;83:45–56. <https://doi.org/10.1101/sqb.2018.83.037887>.
2. Grant SGN. Synapse molecular complexity and the plasticity behaviour problem. *Brain Neurosci Adv.* 2018b. <https://doi.org/10.1177/2398212818810685>.
3. Grant SGN. Synapse diversity and synaptome architecture in human genetic disorders. *Hum Mol Gen.* 2019;28(R2):R219–25. <https://doi.org/10.1093/hmg/ddz178>.
4. Jasinska M, Woznicka O, Jasek-Gajda E, Lis GJ, Pyza E, Litwin JA. Circadian changes of dendritic spine geometry in mouse barrel cortex. *Front Neurosci.* 2020;14:57881. <https://doi.org/10.3389/fnins.2020.578881>.
5. Zeng M, Chen X, Guan D, Xu J, Wu H, Tong P, Zhang M. Reconstituted postsynaptic density as a molecular platform for understanding synapse formation and plasticity. *Cell.* 2018;174:1172–e118716. <https://doi.org/10.1016/j.cell.2018.06.047>.
6. Kalman ZE, Dudola D, Mészáros B, Gáspári Z, Dobson L. PSINDB: the postsynaptic protein–protein interaction database. *Database.* 2022; 2022:baac007. <https://doi.org/10.1093/database/baac007>.
7. Mameza MG, Dvoretzkova E, Bamann M, Hönck H-H, Güler T, Boeckers TM, Schoen M, Verpelli C, Sala C, Barsukov I, Dityatev A, Kreienkamp H-J. SHANK3 gene mutations associated with autism facilitate ligand binding to the shank3 ankyrin repeat region. *J Biol Chem.* 2013;288:26697–708. <https://doi.org/10.1074/jbc.m112.424747>.

8. Kursula P, Shanks — multidomain molecular scaffolds of the postsynaptic density. *Curr Opin Struct Biol.* 2019;54:122–8. <https://doi.org/10.1016/j.sbi.2019.01.007>.
9. Delling JP, Boeckers TM. Comparison of SHANK3 deficiency in animal models: phenotypes, treatment strategies, and translational implications. *J Neurodev Disord.* 2021;13:55. <https://doi.org/10.1186/s11689-021-09397-8>.
10. Leblond CS, Nava C, Polge A, Gauthier J, Huguet G, Lumbroso S, et al. Meta-analysis of SHANK mutations in autism spectrum disorders: a gradient of severity in cognitive impairments. *PLoS Genet.* 2014;10:e1004580. <https://doi.org/10.1371/journal.pgen.1004580>.
11. Schön M, Lapunzina P, Nevado J, Mattina T, Gunnarsson C, Hadziszewski K, et al. Definition and clinical variability of SHANK3-related phelan-McDermid syndrome. *Eur J Med Genet.* 2023;66:104754. <https://doi.org/10.1016/j.ejmg.2023.104754>.
12. Betancur C, Buxbaum JD. SHANK3 haploinsufficiency: a common but underdiagnosed highly penetrant monogenic cause of autism spectrum disorders. *Mol Autism.* 2013;4:17. <https://doi.org/10.1186/2040-2392-4-17>.
13. Sarkar D, Kang J, Wassie AT, Schroeder ME, Peng Z, Tarr TB, et al. Revealing nanostructures in brain tissue via protein decrowding by iterative expansion microscopy. *Nat Biomed Eng.* 2022;6:1057–73. <https://doi.org/10.1038/s41551-022-00912-3>.
14. Ramsey AM, Tang A-H, LeGates TA, Gou X-Y, Carbone BE, Thompson SM, et al. Subsynaptic positioning of AMPARs by LRRTM2 controls synaptic strength. *Sci Adv.* 2021;7:eabf3126. <https://doi.org/10.1126/sciadv.abf3126>.
15. Miski M, Keömley-Horváth BM, Megyeriné DR, Csikász-Nagy A, Gáspári Z. Diversity of synaptic protein complexes as a function of the abundance of their constituent proteins: a modeling approach. *PLoS Comp Biol.* 2022;18:e1009758. <https://doi.org/10.1371/journal.pcbi.1009758>.
16. Ali M, McAuley MM, Lüchow S, Knapp S, Joerger AC, Ivarsson Y. Integrated analysis of Shank1 pdz interactions with c-terminal and internal binding motifs. *Curr Res Struct Biol.* 2021;3:41–50. <https://doi.org/10.1016/j.crstbi.2021.01.001>.
17. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 2018;47:D941–7. <https://doi.org/10.1093/nar/gky1015>.
18. Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature.* 2014;508:199–206. <https://doi.org/10.1038/nature13185>.
19. Reguly Z, Cserecsik D, Juhász J, Tornai K, Bujtár Z, Horváth G, et al. Microsimulation based quantitative analysis of COVID-19 management strategies. *PLoS Comp Biol.* 2022;18:e1009693. <https://doi.org/10.1371/journal.pcbi.1009693>.
20. Rizzetto S, Priami C, Csikász-Nagy A. Qualitative and quantitative protein complex prediction through proteome-wide simulations. *PLoS Comp Biol.* 2015;11:e1004424. <https://doi.org/10.1371/journal.pcbi.1004424>.
21. Liu Y, Beyer A, Aebersold R. On the dependency of cellular protein levels on mRNA abundance. *Cell.* 2016;165:535–50. <https://doi.org/10.1016/j.cell.2016.03.014>.
22. Edfors F, Danielsson F, Hallström BM, Käll L, Lundberg E, Pontén F, Forsström B, Uhlen M. Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol Syst Biol.* 2016;12:883. <https://doi.org/10.15252/msb.20167144>.
23. Feng Z, Chen X, Zeng M, Zhang M. Phase separation as a mechanism for assembling dynamic postsynaptic density signalling complexes. *Curr Opin Neurobiol.* 2019;57:1–8. <https://doi.org/10.1016/j.conb.2018.12.001>.
24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B.* 1995;57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
25. Lee JH, Park H, Park SJ, Kim HJ, Eom SH. The structural flexibility of the Shank1 PDZ domain is important for its binding to different ligands. *Biochem Biophys Res Commun.* 2011;407:207–12. <https://doi.org/10.1016/j.bbrc.2011.02.141>.
26. Giollo M, Martin AJ, Walsh I, Ferrari C, Tosatto SC. NeEMO: a method using residue interaction networks to improve prediction of protein stability upon mutation. *BMC Genomics.* 2014;15(Suppl4):S7. <https://doi.org/10.1186/1471-2164-15-s4-s7>.
27. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental algorithms for Scientific Computing in Python. *Nat Methods.* 2020;17:261–72. <https://doi.org/10.1038/s41592-019-0686-2>.
28. Sudarov A. Defining the role of cerebellar purkinje cells in autism spectrum disorders. *Cerebellum.* 2013;12:950–5. <https://doi.org/10.1007/s12311-013-0490-y>.
29. Wilson HL. Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. *J Med Genet.* 2003;40:575–84. <https://doi.org/10.1136/jmg.40.8.575>.
30. Tsai PT. Autism and cerebellar dysfunction: evidence from animal models. *Semin Fetal Neonatal Med.* 2016;21:349–55. <https://doi.org/10.1016/j.siny.2016.04.009>.
31. Sungur AÖ, Redecker TM, Andres E, Dürichen W, Schwarting RKW, del Rey A, Wöhr M. Reduced efficacy of d-amphetamine and 3, 4-methylenedioxymethamphetamine in inducing hyperactivity in mice lacking the postsynaptic scaffolding protein SHANK1. *Front Mol Neurosci.* 2018;11:419. <https://doi.org/10.3389/fnmol.2018>.
32. Zeng M, Chen X, Guan D, Xu J, Wu H, Tong P, Zhang M. Reconstituted postsynaptic density as a molecular platform for understanding synapse formation and plasticity. *Cell.* 2018;174:1172–87. <https://doi.org/10.1016/j.cell.2018.06.047>.
33. Barti B, Dudok B, Kenesei K, Zöldi M, Miczán V, Balla GY, Zala D, Tasso M, Sagheddu C, Kisfali M, Tóth B, Ledri M, Vizi ES, Melis M, Barna L, Lenkei Z, Soltész I, Katona I. Presynaptic nanoscale components of retrograde synaptic signaling. *Sci Adv.* 2024;10:eado0077. <https://doi.org/10.1126/sciadv.ado0077>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.