Driving forces for condensation of synapsin are governed by sequence-encoded molecular grammars

Christian Hoffmann [a,1], Kiersten M. Ruff [b,1], Irina A. Edu [c], Min Kyung Shinn [b], Johannes V. Tromm [a], Matthew R. King [b], Avnika Pant [b], Hannes Ausserwöger [c], Jennifer R. Morgan [d], Tuomas P. J. Knowles [c,e], Rohit V. Pappu [b,2], Dragomir Milovanovic [a, f, g, h 2]

[a] Laboratory of Molecular Neuroscience Berlin, German Center for Neurodegenerative Diseases (DZNE), 10117 Berlin, Germany.
[b] Department of Biomedical Engineering and Center for Biomolecular Condensates, James McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO, USA.
[c] Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom.
[d] Eugene Bell Center for Regenerative Biology and Tissue Engineering, Marine Biological Laboratory, Woods Hole, MA 02543, USA.
[e] Cavendish Laboratory, Department of Physics, University of Cambridge, JJ Thomson Road, Cambridge CB3 0HE, United Kingdom.
[f] German Center for Neurodegenerative Diseases (DZNE), 53127 Bonn, Germany.
[g] Einstein Center for Neuroscience, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität Berlin, and Berlin Institute of Health, 10117 Berlin, Germany.
[h] Whitman Center, Marine Biological Laboratory, 02543 Woods Hole, MA, USA.

[1]These authors contributed equally
[2]To whom correspondence may be addressed: pappu@wustl.edu; dragomir.milovanovic@dzne.de

Author Contributions: CH performed all reconstitution experiments. KMR did all computational analyses. CH, IE, HA performed microfluidic experiments and analyzed the data; CH and MKS performed cluster distribution in the subsaturated solution measurements and analyzed the data. CH, MK, AP performed pH-based assays and analyzed the data. CH and JVT performed cell experiments. JRM and TPJK contributed to experimental design and discussions. RVP and DM designed the study and wrote the paper. All the authors read and approved the final version of the manuscript.

Competing Interest Statement: RVP is a member of the scientific advisory board and shareholder in Dewpoint Therapeutics Inc. TPJK is the co-founder of Fluid Analytics, Wren Therapeutics, Xampla, and Transition Bio. These affiliations did not contribute to, motivate, or influence the current study. All other authors have no competing interests to declare.

Classification: Major: Biological Sciences/Biochemistry;
Minor: Physical Sciences/Applied Physical Sciences /

Keywords: biomolecular condensates | intrinsically disordered protein regions | synapse

Abstract

Brain functioning relies on orchestrated synaptic vesicle dynamics and controlled neurotransmitter release. Multiple biomolecular condensates coexist at the pre- and post-synapse and they are driven by condensation that combines binding, phase separation, and percolation. In pre-synapses, intrinsically disordered regions (IDRs) of synaptic proteins are drivers of condensation that enable clustering of synaptic vesicles (SVs). Although sequences of IDRs are poorly conserved across evolution, our computational analysis reveals the existence of non-random compositional biases and sequence patterns (molecular grammars) in IDRs of pre-synaptic proteins. For example, synapsin-1, which is essential for condensation of SVs, contains a conserved valence of arginine residues and blocks of polar and proline residues that are segregated from one another along the linear sequence. We show that these conserved features are crucial for driving synapsin-1 condensation *in vitro* and in cells. Our results highlight how conserved molecular grammars drive the condensation of key proteins at the pre-synapse.

Main Text

Functional neurotransmission relies on synaptic vesicle (SV) cycles at the synapse. SVs cluster at pre-synapses forming biomolecular condensates with synapsins, which are abundant synaptic proteins (1)(2). Synapsin-SV condensates enable reversible sequestration and maintenance of soluble synaptic proteins that are essential for distinct steps of SV cycles (3). Extant data suggest that pre-synapses are complex emulsions defined by multiple coexisting phases (4). Here, we report that pre-synaptic proteins feature intrinsically disordered regions (IDRs) with evolutionarily conserved, distinct and complementary molecular grammars. The grammars with synapsin-1 IDR influence condensation via homotypic interactions *in vitro*, and synapsin co-condensates in cells.

We analyzed IDRs of pre-synaptic proteins and uncovered non-random compositional biases and binary sequence patterns (5) – molecular grammars (6) – within synapsin-1 and other SV proteins. Cytosolic IDRs of integral SV proteins are enriched in negatively charged and aromatic residues whereas IDRs of soluble pre-synaptic proteins are enriched in proline and positively charged residues (Fig. 1A-C). Grammars of IDRs of pre-synaptic proteins are distinct from IDRs in proteins from other condensates (8).

The synapsin-1 IDR features segregated blocks of polar/proline residues and a preference for Arg over Lys (Fig. 1D). We designed two variants of synapsin-1 to test how conserved grammars affect condensation. In the SCR variant, we scrambled the IDR sequence to disrupt proline/polar blocks (Fig. 1E). In the RtoK variant, Arg residues within the IDR were substituted with Lys (Fig. 1F). We cloned, expressed, and purified the WT and the two mutants each with an EGFP fluorophore (Fig. 1G,H). WT synapsin-1 readily formed condensates; synapsin-1 SCR formed smaller and more sparse condensates. Synapsin-1 RtoK did not form condensates under equivalent conditions (Fig. 1I) suggesting that substitution of Arg to Lys weakens homotypic interactions (9).

Next, we co-expressed each synapsin-1 construct in cells with synaptophysin, an integral SV protein implicated in the formation of small vesicles (10). The IDR of synaptophysin is enriched in aromatic residues (Fig. 1B). Co-expression of WT synapsin-1 recapitulated co-condensation with synaptophysin (11). However, co-expression of synapsin-1 SCR

with synaptophysin condensates generated a gain-of-function phenotype for cytoskeleton binding without forming condensates. Co-expressed synapsin-1 RtoK remained fully soluble (Fig. 1J). Substitution of Arg with Lys in the synapsin-1 IDR abrogated co-condensation with synaptophysin in cells.

For rigorous quantification of driving forces, we mapped comparative phase boundaries for the synapsin-1 variants using a microfluidics-based "PhaseScan" approach (12). Synapsin-1 WT, synapsin-1 SCR and synapsin-1 RtoK showed distinct phase boundaries (Fig. 2A) with the synapsin-1 RtoK being the weakest driver of phase separation (Fig. 2B). We also deployed microfluidics resistive pulse sensing (MRPS) to quantify size distributions of clusters in subsaturated solutions (13). Substitution of Arg to Lys alters the size distribution showing a pronounced shift toward smaller species (Fig. 2C,D). Even though the driving forces for phase separation are weakened when the IDR is scrambled, clustering in subsaturated solutions is equivalent for synapsin-1 WT and synapsin-1 SCR. This highlights how driving forces for forming pre-percolation clusters and phase separation can be decoupled (13).

Motivated by recent work on nucleoli (8), we asked if phase separation generates a pH gradient because this could help with proper loading of neurotransmitters into SVs (14) (15). Using the ratiometric pH-sensitive dye SNARF-1 (Fig. 2E-2G), we found that synapsin-1 WT, SCR, and RtoK showed equivalent abilities to accumulate protons in dense phases (Fig. 2H). Therefore, while molecular grammars in synapsin-1 control driving forces for condensation, the interphase pH gradients remain unchanged, presumably because sequences have identical net charge (8).

Our work shows how insights from computational analysis can be tested via designed mutations. This approach bypasses the often-used approach of deleting IDRs to assess their contributions – an approach that rests on the erroneous notion that IDRs engage mainly in non-specific interactions. Instead, we show how systematic designs that modulate conserved grammars can be brought to bear to investigate how these regions contribute to phase behavior. In the context of the pre-synapse, the results presented here and investigations of the synergies of molecular grammars in synaptic boutons (4) are likely to enable new insights into our understanding of the neuronal transmission that is relevant in healthy brains and in diseases characterized by the loss of functional synapses.

Materials and Methods. Details of the materials and methods are available in Supplementary Information.

Acknowledgments

We thank the Advanced Medical Bioimaging Core Facility at Charité University Clinic, Berlin for microscopy. This work was supported by DZNE, grants from the DFG (MI 2104 and SFB1286/B10 to DM); ERC Grants (MemlessInterface 101078172 to DM, DiProPhys 101001615 to TPJK); the NIH (NIA 2RF1 NS078165-12 to JRM, NINDS R01NS121114 to RVP, and F32GM146418-01A1 to MRK) and AFOSR (FA9550-20-1-0241 to RVP). CH is supported by the Innovative Minds Program of the German Dementia Association.

Figures

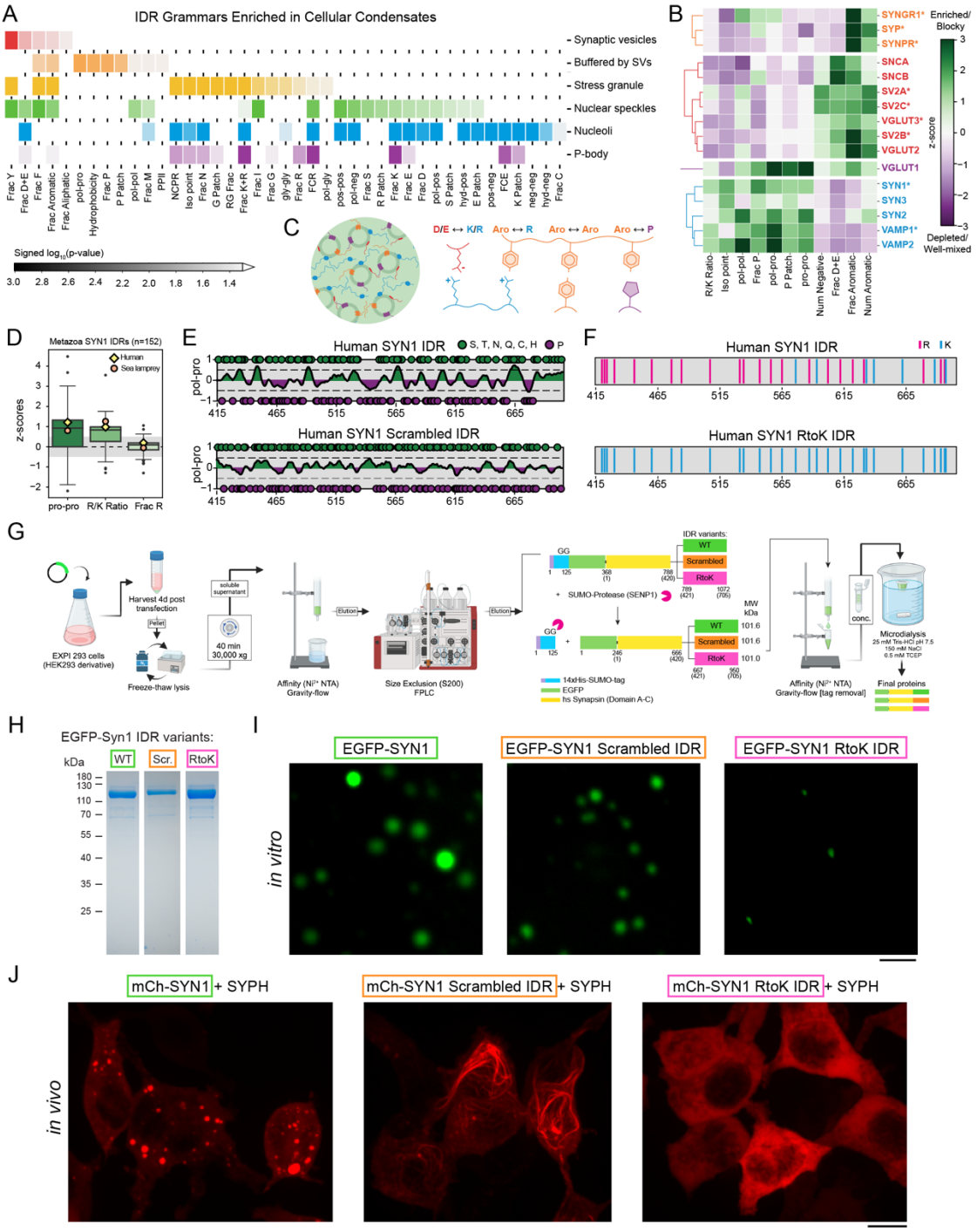

Figure 1: IDRs of proteins that make up synaptic condensates feature distinct molecular grammars.

A. IDR grammars of SV proteins are distinct from those of IDRs in other condensates.

B. IDR compositions and sequence patterns of pre-synaptic proteins fall into distinct clusters.

C. Schematic illustrating the chemical complementarity across IDRs within SV clusters.
D. Molecular grammars are conserved across 152 metazoans, including lamprey, which is the oldest vertebrate predecessor of humans.
E. Schematic showing how the synapsin-1 scrambled IDR in which polar and proline blocks are evenly distributed to reduce blockiness.
F. Schematic of the synapsin-1 IDR showing the substitution of Arg with Lys.
G. Purification protocol for synapsin-1 IDR (WT) and the two variants SCR and RtoK.
H. Coomassie gel indicating the final recombinant proteins used for the biophysical assays.
I. Microscopy images of EGFP-synapsin-1 (8 μM) with WT, SCR, and RtoK IDR reconstituted in 25 mM Tris-HCl (pH 7.5), 0.5 mM TCEP, 150 mM NaCl and 3% PEG 8,000. Scale bar, 5 μm.
J. Images of HEK cells expressing EGFP-synapsin-1 WT (left), EGFP-synapsin-1 scrambled IDR (middle), or EGFP-synapsin-1 RtoK IDR (right) with untagged synaptophysin. Mutants disrupted formation of normal synapsin condensates. IDR scramble formed filaments resembling cytoskeleton, and RtoK completely abolished condensation. Scale bar, 10 μm.

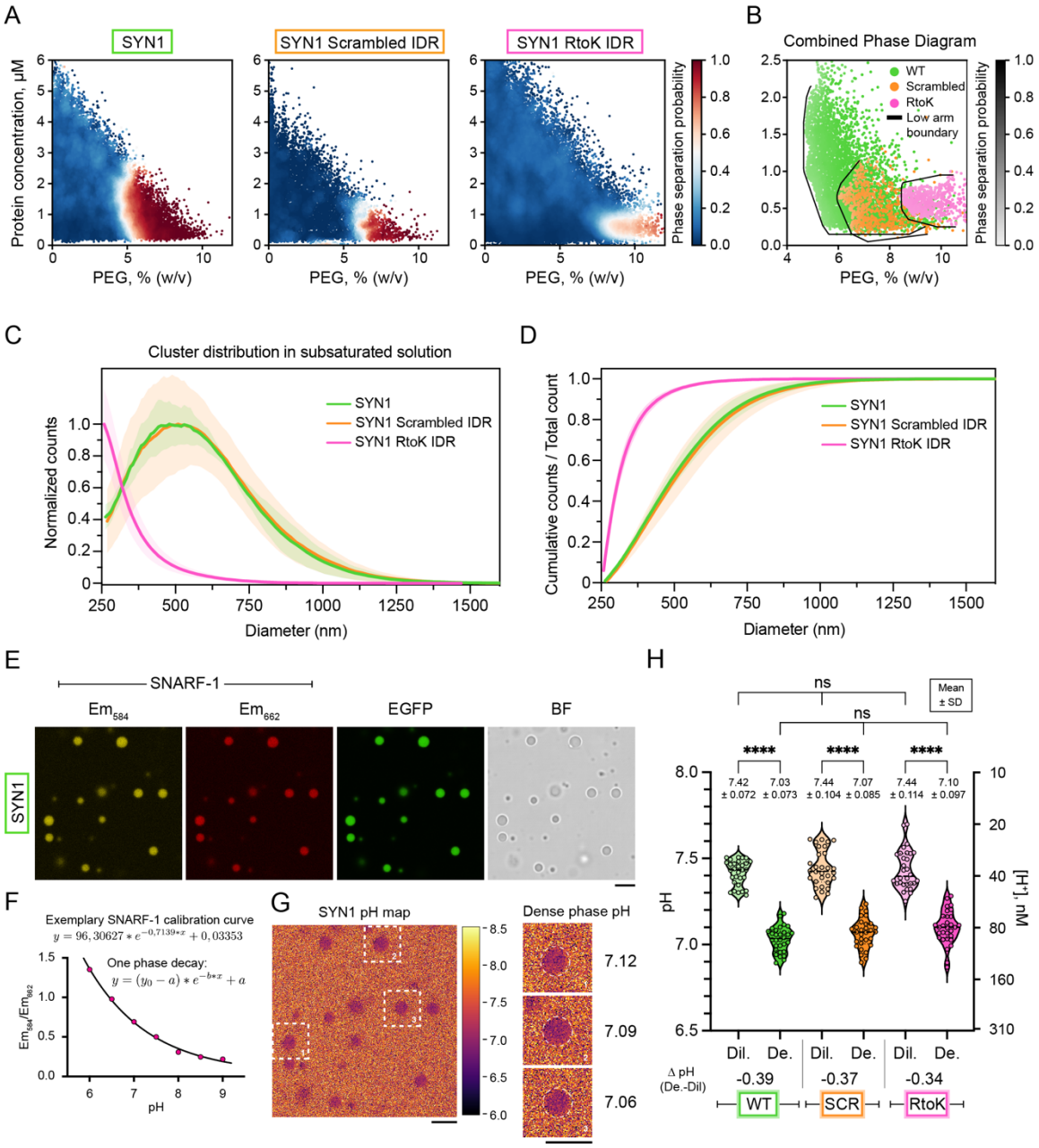

Figure 2. Molecular grammars of synapsin-1 IDR influence synapsin-1 condensation.

A. Low concentration arms of phase boundaries of EGFP-synapsin-1 WT (left, N = 29,187 points), EGFP-synapsin-1 scrambled IDR (middle, N = 13,203 data points), or EGFP-synapsin-1 RtoK IDR (right, N = 64,675 data points). Red and blue data points correspond to individual microdroplets classified as phase separated or homogeneous, respectively. The heatmap quantifies the phase separation probability.

B. Combined phase boundaries from **A** showing the low concentration arm.

C. Cluster size distributions in subsaturated solutions for synapsin-1 WT (green), SCR (orange), and RtoK (magenta) variants. Measurements were performed at equivalent

conditions (8 µM Synapsin 1 variants with 3% PEG 8,000 in 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM TCEP) using a C-2000 cartridge in an nCS1 instrument. The plot shows counts that are normalized to the maximum value as a function of the particle diameter.

D. Cumulative counts for cluster size distributions shown in *C.* Error bars (shaded) represent standard deviation.

E. Representative micrographs of EGFP-synapsin-1 condensates (8 µM protein in 3% PEG 8,000) co-incubated with the ratiometric pH-sensitive reporter SNARF-1. SNARF-1 was excited at 561 nm and recorded at 582-586 (yellow) and 660-664 nm (red). Scale bar, 5 µm.

F. Carboxy SNARF-1 calibration curve for reconstitution buffer (25 mM Tris-HCl, 150 mM NaCl, 0.5 mM TCEP) at different pH (pH 6 – pH 9) a single plane. Data were fit to a one-phase-decay function.

G. Left: Example micrograph indicating a pH map for EGFP-synapsin-1 WT condensates calculated by applying the z-slice calibration curve. Heatmap represents pH. Right: Magnified regions of selected condensates (dashed line) with indicated average pH values of the dense phase. Scale bar, 5 µm.

H. pH analysis for the dense and dilute phases of condensates for synapsin-1 WT (green), SCR (orange), and RtoK (magenta) variants. Average pH values are shown in the graph with standard deviation. Violin plots with median and quartiles indicated by the dashed lines. (One-way ANOVA, ****: $p<0,0001$).

References

1. D. Milovanovic, Y. Wu, X. Bian, P. De Camilli, A liquid phase of synapsin and lipid vesicles. *Science* **361**, 604 607 (2018).

2. C. Hoffmann, *et al.*, Synapsin condensation controls synaptic vesicle sequestering and dynamics. *Nature Commun* **14**, 6730 (2023).

3. A. Denker, K. Kröhnert, J. Bückers, E. Neher, S. O. Rizzoli, The reserve pool of synaptic vesicles acts as a buffer for proteins involved in synaptic vesicle recycling. *Proc Natl Acad Sci USA* **108**, 17183 17188 (2011).

4. R. Sansevrino, C. Hoffmann, D. Milovanovic, Condensate biology of synaptic vesicle clusters. *Trends Neurosci.* **46**, 293-306 (2023).

5. M.C. Cohan, *et al.*, Uncovering non-random binary patterns within sequences of intrinsically disordered proteins. *J Mol Biol* **434**, 167373 (2022).

6. J. Wang, *et al.*, A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699.e16 (2018).

7. Z. Taoufiq, *et al.*, Hidden proteome of synaptic vesicles in the mammalian brain. *Proc Natl Acad Sci USA* **117**, 33586–33596 (2020).

8. M. R. King, *et al.*, Macromolecular condensation organizes nucleolar sub-phases to set up a pH gradient. *Cell* **187**, 1889-1906.e24 (2024).

9. R. S. Fisher, S. Elbaum-Garfinkle, Tunable multiphase dynamics of arginine and lysine liquid condensates. *Nature Commun* **11**, 4628 (2020).

10. P. L. Cameron, T. C. Südhof, R. Jahn, P. De Camilli, Colocalization of synaptophysin with transferrin receptors: implications for synaptic vesicle biogenesis. *J Cell Biol* **115**, 151–164 (1991).

11. D. Park, *et al.*, Cooperative function of synaptophysin and synapsin in the generation of synaptic vesicle-like clusters in non-neuronal cells. *Nature Commun.* **12**, 263 (2021).

12. W. E. Arter, *et al.*, Biomolecular condensate phase diagrams with a combinatorial microdroplet platform. *Nature Commun* **13**, 7845 (2022).

13. M. Kar, *et al.*, Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proc Natl Acad Sci USA* **119**, e2202222119 (2022).

14. Z. Farsi, *et al.*, Single-vesicle imaging reveals different transport mechanisms between glutamatergic and GABAergic vesicles. *Science* **351**, 981–984 (2016).

15. E. Kosmidis, *et al.*, Regulation of the mammalian-brain V-ATPase through ultraslow mode-switching. *Nature* **611**, 827–834 (2022).

Supporting Information for

Driving forces for condensation of synapsin are governed by sequence-encoded molecular grammars

Christian Hoffmann [a,1], Kiersten M. Ruff [b,1], Irina Edu [c], Min Kyung Shinn [b], Johannes V. Tromm [a], Matthew R. King [b], Avnika Pant [b], Hannes Ausserwöger [c], Jennifer R. Morgan [d], Tuomas P. J. Knowles [c,e], Rohit V. Pappu [b,2], Dragomir Milovanovic [a, f, g, h, 2]

This PDF file includes:
Extended Methods
SI References

Supporting Information Text

Analysis of IDRs using NARDINI+

Sequence grammar analysis was performed using NARDINI+ as described in (1). Briefly, NARDINI+ calculates z-scores of compositional and patterning grammar features previously shown to be important for IDR conformation, interactions, and function (2). Five additional compositional grammar features were examined that were not described in King et al. These include the number of positive, negative, polar, aliphatic, and aromatic residues, giving a readout on valence compared to the full IDRome. For a given IDR, compositional z-scores are calculated by extracting the compositional feature of the given IDR and using the mean and standard deviations from the full human IDRome (24508 IDRs). The human IDRome contains all IDRs of length ≥ 30, extracted from the UniProt human proteome (UP000005640) and MobiDB predictions (3) To calculate the patterning z-scores for a given IDR, NARDINI is performed (4). Here, the asymmetry of patterning of two residue types in a given IDR sequence is compared to the mean and standard deviations of 10^5 random sequences of the same composition.

Extraction of condensate-specific enriched grammar features

Condensate specific protein lists were extracted from various sources and cross-referenced with the human IDRome to extract condensate specific IDRs. As described in (1), proteins localized to nucleoli and nuclear speckles were extracted from the Human Protein Atlas (HPA). Proteins localized to P-bodies and stress granules were extracted from (5) and (6), respectively. Synaptic vesicle proteins were extracted from (7) using the list of “SV-resident” proteins leading to 125 IDR sequences. Buffered by SV proteins were extracted from (8) (here paralogs were included) to yield 47 IDRs. To determine which grammar features were enriched in specific condensates, the human IDRome was split into two sets: (1) the current condensate-specific IDR set and (2) the remaining human IDRome. Then, the distributions for each of the 90 different z-scores were compared using the two-sample Kolmogorov-Smirnov test and a p-value was calculated. The signed

$\log_{10}$(p-value) was calculated for p-values < 0.05. Specifically, -$\log_{10}$(p-value) was calculated if the condensate-specific mean z-score value was greater than the rest of the IDRome mean z-score value and $\log_{10}$(p-value) was calculated if the condensate-specific mean z-score was smaller. Thus, positive values imply the given feature is enriched or blockier in the condensate-specific IDR set.

Extraction of IDR grammar clusters within SV proteins

Grammar feature z-scores were calculated using NARDINI+ as described above for IDRs of SV proteins: SNCA, SNCB, SNCG, SYN1, SYN2, SYN3, SYP, SYNPR, SYNGR1, SV2A, SV2B, SV2C, STX7, VAMP1, VAMP2, VGLUT1, VGLUT2, and VGLUT3. Figure 1B shows a filtered set of enriched features for the four clusters having at least 3 IDRs determined by hierarchical clustering using the Euclidean distance and Ward's linkage method (9).

Evolutionary analysis of Synapsin IDRs

The list of eukaryotic orthologous SYN1 orthologs were extracted from EggNOG (10). The proteomes of 12 additional species were examined for orthologs if not included in the orginal EggNOG list. These include *Aptenodytes forsteri* (taxid: 9233), *Doryteuthis pealeii* (taxid: 1051067), *Lampetra fluviatilis* (taxid: 7748), *Leptonychotes weddellii* (taxid: 9713), *Marmota marmota marmota* (taxid: 9994), *Microcebus murinus* (taxid: 30608), *Octopus vulgaris* (taxid: 6645), *Petromyzon marinus* (taxid: 7757), *Physeter macrocephalus* (taxid: 9755), *Urocitellus parryii* (taxid: 9999), Ursus americanus (taxid: 9643), and *Ziphius cavirostris* (taxid: 9760) whose proteomes were extracted from UniProt. BLASTp was used to extract only the most similar ortholog per species as compared to the full SYN1 sequence. Then, MUSCLE sequence alignment was utilized to align orthologous sequences (11). From the alignments, Jalview was used to extract the IDR regions in orthologs corresponding to residues 415-705 of human SYN1 (12). Only orthologous sequences of length ≥ 15 and housing no unnatural amino acids were kept. NARDINI+ was then ran on all orthologous sequences where the human IDRome was used as the prior to extract the z-scores of the compositional features. Overall, 152 species were analyzed.

Calculation of mean net residue type profiles

Here, we consider residue types pol≡{C, H, N, Q, S, T} and pro≡P. For a given sequence, the fraction of pol residues minus the fraction of pro residues is calculated for each sliding window of length five. Then, the values from all sliding windows that contain a given residue are averaged to yield a residue specific mean net residue type value. These values are plotted in Figure 1D.

Cloning of Synapsin 1 IDR variants

Human synapsin-1 coding sequence is based on the NM_006950.3 NCBI reference Sequence. The complexity of the sequence, such as GC-rich regions, was manually altered to allow efficient gene synthesis while leaving the protein sequence unchanged (NP_008881.2). An endogenous SacI restriction site at codons E408-E409 (GAGCTC) was silenced by altering the nucleotide sequence to GAGCTG. To allow for easy subsequent exchange of the synapsin-1 IDR, a BamHI was added at codons G389-S390 by changing the nucleotide sequence from GGTTCC to GGATCC. The optimized human synapsin-1 codon sequence and the variant IDR sequences (a.a. 389-705) were created by gene synthesis (Eurofins, HIFI DNA assembly). Subsequently, IDR fragments

starting from a.a. 389 were subcloned using the restriction sites BamHI and SacI, yielding sequences of full-length Synapsin 1 containing IDR variants. The resulting full-length synapsin-1 sequences (WT, scrambled IDR, RtoK IDR) were either subcloned into a pCMV expression plasmid as a C-terminal fusion protein to mCherry for microscopy (13) by Kpn2I and SacI or subcloned into a His14-SUMO_Eu1-EGFP-MCS plasmid for protein purification by BglII and KpnI (14). All constructs were verified by Sanger sequencing.

Cell culture and transfection

HEK 293T cells were grown in Dulbecco's Modified Eagle's Medium GlutaMAX supplement (10566016, Gibco) supplemented with 10% fetal bovine serum (12106C, Sigma), 1% MEM Non-essential Amino Acid Solution (M7145, Sigma) and 1% penicillin/streptomycin (15140122, Gibco) at 37 °C and 5% CO_2. For imaging, cells were seeded on 25 mm cover glasses. Cells were transfected with a total of 2 µg of plasmid DNA using Lipofectamine 2000 (11668500, Invitrogen) following the manufacturer's instructions. In brief, the final transfection mixture contained 3 µL of lipofectamine with 2 µg of total plasmid DNA in 200 µL OptiMEM (31985070, Gibco). Transfection mix was incubated for 20 min at room temperature before adding to seeded HEK cells (confluency ~ 60-70%) and incubated overnight at 37 °C and 5% CO_2. For protein purification experiments, Expi293F™ suspension cultures (A14527, Gibco) were maintained in Expi293™ Expression Medium (A1435102, Gibco) according to the manufacturer's protocol (37 °C, 8% CO2, 125 rpm). Transfection of suspension cultures (30 mL) was performed with 30 µg of plasmid DNA following the ExpiFectamine™ 293 Transfection Kit guidelines.

Protein expression of Synapsin 1 IDR variants

EGFP-tagged synapsin-1 IDR variants were expressed in EXPI cells (HEK derivative) and purified by a two-step purification strategy consisting of NiNTA-affinity and size-exclusion chromatography followed by His-SUMO-tag removal, as described (14) (15). Briefly, His14-SUMO_Eu1-synapsin-1 IDR variants were expressed in Expi293F™cells (Thermo Fisher Scientific) for 3 days following enhancement. Cells were lysed by three cycles of freezing and thawing in a buffer that contained 25 mM Tris-HCl (pH 7.4), 300 mM NaCl, 0.5 mM TCEP (buffer A) supplemented with EDTA-free Roche Complete protease inhibitors, 15 mM imidazole, 10 µg/mL DNaseI and 1 mM $MgCl_2$. The purification steps were carried out at 4°C. Lysate clearance was done by centrifugation for 40 min at 30,000xg. The soluble supernatant was incubated with complete His-Tag Purification resin (Roche) in a polyprep column (Biorad) on a rotating platform for 1 h. Washing steps were done with buffer A and elution was performed in buffer A with 400 mM imidazole. Fractions were concentrated using a 30 K MWCO protein concentrator (Millipore) and applied to size exclusion chromatography (Superdex™200 Increase 10/300, GE Healthcare) in buffer A. Fractions that contained the His14-SUMO_Eu1-Synapsin 1 IDR variant were combined subjected to tag removal with SENP_EuB SUMO protease digest overnight (protease:protein ratio of 1:20). The cleaved His-SUMO-tag and His-tagged SENP protease was removed by supplementing the reaction with 15 mM imidazole and passing it 3 times over preequilibrated (buffer A) complete His-Tag Purification resin. Proteins in the flow-through were concentrated with a 30 K MWCO protein concentrator (Millipore) and subjected to dialysis against 25 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.5 mM TCEP in Thermo Slide-A-Lyzer floating cups (MWCO 10 kDa). Proteins were snap-frozen in liquid nitrogen and stored at -80 °C until use.

Microscopy and image analysis

For *in vitro* reconstitution of EGFP-Synapsin 1 IDR variants, 8 μM protein was coincubated with polyethylene glycol (PEG 8,000) at a final concentration of 3% in a buffer containing 25 mM Tris-HCl (pH 7.4), 150 mM NaCl and 0.5 mM TCEP. After the addition of PEG, the condensation reaction was transferred to a glass bottom dish (Cellvis D35-20-1.5-N). Imaging was performed on a Nikon spinning disk confocal CSU-X microscope equipped with a Plan Apo λ 60x Oil objective and a pco.edge camera. The excitation wavelength was 488 nm for EGFP, and image analysis was performed using ImageJ2 (Version: 2.9.0/1.53t).

For live-cell confocal imaging in HEK cells an Andor DU-888 X-9798 camera and excitation wavelength at 561 nm were used of mCherry-tagged synapsin-1 IDR variants.

The PhaseScan approach for mapping phase diagrams

The microfluidic devices were fabricated using standard soft-lithography (16) (17). Devices were designed using the AutoCAD (AutoDesk) software and then printed on a photomask (Micro Lithography). To obtain a master with microchannels of approximately 50 μm height, SU-8 3050 (Kayaku Advances Materials) was poured onto a silicon wafer (MicroChem) and spun using a spincoater (Laurell Technologies) for 45 s at 3000 rpm. After soft-baking at 95 °C for 25 min, the wafer with the photomask placed on it was subjected to UV exposure for 60 s. Post-exposure baking at 95 °C for 5 min was then followed by developing with propylene glycol methyl ether acetate (PGMEA, Sigma-Aldrich) and a washing step with isopropanol (IPA, VWR Chemicals). To fabricate the devices, polydimethylsiloxane (PDMS, Sylgard 184) was poured over the master, degassed and then baked at 65 °C for 1.5 h. Subsequently, devices were cut out and corresponding holes for the inlets and outlets were punched. A plasma oven (Diener Electronic) was used to treat the devices and enable bonding them to glass slides. The last step was to apply hydrophobic surface modification to the devices using 2% (v/v) trichloro(1H,1H,2H,2H-perfluorooctyl)silane (Sigma-Aldrich) in hydrofluoroether HFE-7500 (Fluorochem).

The PhaseScan microfluidic platform was used to map high-density phase diagrams (16). The following three aqueous solutions were prepared to be mixed on-chip: 6 μM stock of one of the three EGFP labelled synapsin-1 variants in 25 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.5 mM TCEP; solution of the corresponding buffer and a solution of polyethylene glycol (PEG 8000) 10% (w/v) in the same buffer, barcoded with 2 μM Alexa Fluor 647 free dye. Additionally, an oil solution of HFE-7500 (Fluorochem) with 4% surfactant (RAN Biotechnologies) was used for water-in-oil droplet encapsulation. The flow rates of the three aqueous solutions were varied automatically using pressure-controlled pumps (LineUp Flow EZ, Fluigent) according to a pre-programmed flow profile such that each microfluidic droplet contained different concentrations of protein and PEG. The total flow rate of the aqueous solutions was kept constant at 60 μL/h, while the oil flow rate was 150 μL/h, generating droplets of approximately 70 μm in diameter. Droplets were incubated for 3 minutes before imaging in the same microfluidic chip at two wavelengths (488 nm and 647 nm) at the same time using an epifluorescence microscope (Cairn Research), equipped with a 10x air objective (Nikon CFI Plan Fluor) and a dichroic filter set (Cairn Research). A semi-automated custom-written Python script (Python version 3.9.7) was used to analyse the images, i.e. to detect droplets and classify them as phase separated (1) or mixed (0) by employing a convolution neural network (CNN) (18). To allow calculation of the concentration of the components in each droplet, images of calibration droplets containing only the stock solutions of protein or PEG were recorded. A double-

gaussian fit was applied over the fluorescence intensity histograms of the calibration droplets to correlate the intensities back to the stock concentration (19). Hence, each microfluidic droplet provides information for one data point on the phase diagrams presented as scattered plots. Phase separation probability was further determined for each data point by averaging over its neighboring points within a radius of 5% of the maximum limit of each axis. The boundaries were determined using fitting that is based on support vector machine-based methods.

Extracting low concentration arms of phase boundaries

First, we generate a two-dimensional histogram of all points that have a phase separation probability of at least 0.55. To eliminate outliers, if there exists at least three data points for the scramble or RtoK sequences and five data points for the WT sequence in a given bin, then the x and y positions of that bin are kept. The boundary is then drawn using these x and y positions and calculating the convex hull with rounded corners.

Measurements of cluster distribution in the subsaturated solution

Microfluidic Resistive Pulse Sensing (MRPS) measurements were performed on an nCS1 instrument (Spectradyne LLC, Signal Hill, CA, USA). 8 μM of purified synapsin construct was mixed with 3% PEG 8,000 in buffer (25 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.5 mM TCEP) in an Eppendorf tube to induce phase separation. 5 μL of samples were measured using C-2000 cartridges, with a measurement range of 250–2000 nm. At least three acquisitions (triplicates) were collected and combined for analysis, with each acquisition being collected over 10 minutes. Control experiments were performed with 3% PEG 8,000, where no significant abundance of particles was detected. The particle size distributions from the triplicate experiments were averaged and the resulting curve was smoothed using the LOWESS method in Prism 10 (GraphPad, Boston, MA, USA).

Analysis of pH in Synapsin 1 condensates

The ratiometric pH-sensitive dye, 5-(and-6)-Carboxy SNARF™-1 (C1270, Invitrogen), was dissolved to a final concentration of 10 mM in ddH_2O. For calibration, 25 mM Tris-HCl, 150 mM NaCl, and 0.5 mM TCEP buffer were prepared at different pH values (pH 6.0-9.0, steps: 0.5 pH units). All calibration curves were taken on the same day as the samples with synapsin-1 condensate were analyzed (all in triplicates). In brief, the calibration buffer was mixed with 20 μM final SNARF-1 dye (5 μL final reaction volume) and placed on a glass bottom dish (Cellvis D35-20-1.5-N). Imaging was performed on a Nikon laser scanning confocal AX NSPARC microscope (Ti2, AX camera, Galvano scanner, 2x line averaging, 1024x1024 px resolution, 1.0 μs dwell time) equipped with a Plan Apo λD 60x oil OFN25 DIC N2 objective. Ratiometric imaging of the SNARF-1 dye was performed by excitation at 561 nm and simultaneous readout of the two GaAsP PTMs with different freely tunable emission bands. The first band was set to 582-586 nm (center wavelength: 584 nm) and the second band was set to 660-664 nm (center wavelength: 662 nm). A total of 20 z-stacks were acquired from the cover glass surface into the buffer solution with a step size of 0.20 μm (total measured relative z-height 4.0 μm). Over the entire pH range, the calibration curves were acquired for each z-stack at the same distance from the surface of the coverglass and fitted to a one-phase decay function using GraphPad Prism 10.

For the pH analysis of EGFP-synapsin-1 IDR variants, 8 μM protein was coincubated with 3% PEG 8,000 and 20 μM SNARF-1 dye in a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl and 0.5 mM TCEP. Ratiometric z-stacks were recorded

from the cover glass surface with the same step size and ratiometric imaging settings as for the calibration curves. Transmission detection was acquired using the 561-nm laser line. Additionally, imaging of EGFP was performed by excitation at 488 nm and emission at 499-551 nm. Of note, while synapsin-1 WT and scrambled IDR condensates formed readily within minutes, synapsin-1 RtoK IDR variant condensate formation was forced by extending the incubation time on the cover glass to one hour.

For image analysis, each pixel of the 584-nm emission image was divided by the corresponding pixel of the 662-nm emission image for each individual z-stack. The resulting ratio z-stack series was converted to a pH stack by applying the corresponding z-slice calibration curve to each matching ratio z-slice. All measurements were performed in three independent reconstitutions with 10 condensates selected from each single condition. A dense phase ROI was outlined with the segmented line tool in ImageJ, and a corresponding circular dilute phase ROI (diameter: 4.3 µm) was placed 2 µm away from the dense phase. ROIs were selected on a single z-plane. Mean pH values of both dense and dilute phases were plotted as violin plots. One-Way ANOVA tests (multiple comparisons, Tukey, alpha=0.05) were performed for statistical analysis.

SI References

1. M. R. King, *et al.*, Macromolecular condensation organizes nucleolar sub-phases to set up a pH gradient. *Cell* **187**, 1889-1906.e24 (2024).

2. T. Zarin, *et al.*, Identifying molecular features that are associated with biological function of intrinsically disordered protein regions. *eLife* **10**, e60220 (2021).

3. D. Piovesan, *et al.*, MobiDB: 10 years of intrinsically disordered proteins. *Nucleic Acids Res* **51**, D438–D444 (2022).

4. M. C. Cohan, M. K. Shinn, J. M. Lalmansingh, R. V. Pappu, Uncovering Non-random Binary Patterns Within Sequences of Intrinsically Disordered Proteins. *J Mol Biol.* **434**, 167373 (2022).

5. A. Hubstenberger, *et al.*, P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Mol Cell* **68**, 144-157.e5 (2017).

6. S. Jain, *et al.*, ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164**, 487–498 (2016).

7. Z. Taoufiq, *et al.*, Hidden proteome of synaptic vesicles in the mammalian brain. *Proc Natl Acad Sci USA* **117**, 33586–33596 (2020).

8. A. Denker, K. Kröhnert, J. Bückers, E. Neher, S. O. Rizzoli, The reserve pool of synaptic vesicles acts as a buffer for proteins involved in synaptic vesicle recycling. *Proc Natl Acad Sci USA* **108**, 17183 17188 (2011).

9. J. H. Ward, Hierarchical Grouping to Optimize an Objective Function. *J Am Stat Assoc* **58**, 236 (1963).

10. J. Huerta-Cepas, *et al.*, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**, D309–D314 (2019).

11. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).

12. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

13. C. Hoffmann, *et al.*, Synapsin Condensates Recruit alpha-Synuclein. *J Mol Biol* **433:** 166961 (2021).

14. C. Hoffmann, *et al.*, Synapsin condensation controls synaptic vesicle sequestering and dynamics. *Nat Commun* **14**, 6730 (2023).

15. C. Akshita, *et al.*, Condensates of synaptic vesicles and synapsin are molecular beacons for actin sequestering and polymerization. *bioRxiv* 2024.07.19.604346 (2024). https://doi.org/10.1101/2024.07.19.604346.

16. W. E. Arter, *et al.*, Biomolecular condensate phase diagrams with a combinatorial microdroplet platform. *Nat Commun* **13**, 7845 (2022).

17. D. Qin, Y. Xia, G. M. Whitesides, Soft lithography for micro- and nanoscale patterning. *Nat Protoc* **5**, 491–502 (2010).

18. N. A. Erkamp, *et al.*, Spatially non-uniform condensates emerge from dynamically arrested phase separation. *Nat Commun* **14**, 684 (2023).

19. H. Ausserwöger, *et al.*, Biomolecular condensates sustain pH gradients at equilibrium driven by charge neutralisation. *bioRxiv* 2024.05.23.595321 (2024). https://doi.org/10.1101/2024.05.23.595321.