Supplemental material

Proteomic screening of glutamatergic mouse brain synaptosomes isolated by fluorescence activated sorting

Christoph Biesemann, Mads Grønborg, Elisa Luquet, Sven P. Wichert, Véronique Bernard, Simon R. Bungers, Ben Cooper, Frédérique Varoqueaux, Liyi Li, Jennifer A. Byrne, Henning Urlaub, Olaf Jahn, Nils Brose, and Etienne Herzog

Online supplemental results

Figure S1 shows a multiple component model of the fluorescence intensity distributions of the flow cytometry data of FASS samples and unsorted synaptosomes. This model serves to more accurately estimate the enrichment of VGLUT1$^{VENUS}$-positive synaptosomes in the FASS samples.

Figure S2 shows the meta-analysis of our proteomics data and the transcriptomics data from (Cahoy et al., 2008). We compared most ≥ 2-fold enriched or depleted proteins, with cell type-specific mRNA expression data, and clustered them accordingly. In contrast, protein products of genes with maximal mRNA expression in astrocytes were 4-fold more abundant in the FASS-depleted fraction.

Figure S3 shows complementary data on the characterization of FXYD6 expression in cultured neurons and the brain, in order to further substantiate our findings that FXYD6 is localized to synapses.

Figure S4 shows complementary data on the characterization of Tpd52 expression in dissociated neuron cultures. Tpd52 is shown to colocalize with VGLUT1 but not with GAD65, and to partially colocalize with the dendritic marker MAP2.

Table S1 lists proteins identified in the Mascot search using MS data on unsorted synaptosomes.
Table S2 lists proteins identified in the Mascot search using MS data on sorted VGLUT1\textsuperscript{VENUS}-positive synaptosomes.

Table S3 lists the complete results of our spectral counting analysis using Scaffold. Furthermore, this table also integrates cell type-specific transcriptome data (Cahoy et al, 2008) with our spectral counting analysis. All proteins analyzed by Western blotting or immunofluorescence staining in our study are indicated.

Table S4 classifies the 163 proteins enriched 2-fold or more in FASS samples according to subcellular localization and cellular function categories.

**Supplemental methods**

**Subcellular fractionation**
The subcellular fractionation was adapted from different published protocols (Hebb & Whittaker, 1958; Huttner et al, 1983; Jones & Matus, 1974) and essentially followed the same basic protocol as used for preparation of synaptosomes for FASS (see above). All steps were performed at 4°C in the presence of 4 mM HEPES pH7.4 and protease inhibitors (1 µM PMSF, 1 µg/ml Aprotinin, 1 µg/ml Leupeptin). The P2 pellet was carefully resuspended in 8 ml of homogenization buffer and transferred to a new centrifuge tube, avoiding the dark brown mitochondria-rich part of the P2 pellet, and centrifuged once more at 12,500 x g for 15 min at 4°C in an SS-34 rotor (Sorvall). The resulting supernatant (S2\textsubscript{1}) was combined with the first S2 fraction, mixed, and centrifuged at 223,160 x g for 2 h at 4°C using a 75Ti fixed angle rotor (Beckman). This centrifugation yielded the microsomal fraction (P3) and the cytosolic supernatant (S3). The final P2 fraction was resuspended in 500 µl of homogenization buffer. Hypotonic shock was used to release of the contents of P2 particles by addition of 9 ml of ice cold lysis buffer (H\textsubscript{2}O, 4 mM HEPES pH 7.4, protease inhibitors as indicated above) and homogenization in a glass-Teflon homogenizer (1,500 rpm; three strokes). The lysed P2 was centrifuged at 25,000 x g for 20 min at 4°C in an SS-34 rotor (Sorvall), resulting in the 'lysed pellet 1' (LP1) and the 'lysed supernatant 1' (LS1). The LS1 fraction was centrifuged at 223,160 x g for 2 h at 4°C in a 75Ti fixed angle rotor (Beckman), resulting in the 'lysed
supernatant 2' (LS2) and the 'lysed pellet 2' (LP2). The LS1 fraction is enriched in soluble synaptic proteins and was concentrated in Centricon 10 cartridges (Millipore). The LP1 fraction was resuspended in 1 ml of homogenization buffer and applied to a sucrose gradient as for synaptosome preparation. The sample was then centrifuged at 64,206 x g (SW41Ti rotor, Beckman) for 2 h at 4°C. The interface at 0.8 M and 1.2 M sucrose contained the synaptic plasma membrane fraction (LP1B). The less dense myelin rich membrane fraction (LP1A) was collected from the interface at 0.32 M/0.8 M sucrose concentrations.

**Multiple normal distribution component model**

For further analyses, flow cytometry data were exported as FCS3.0 files. Raw data were extracted from FCS3.0 files using FCSextract (http://research.stowers-institute.org/efg/ScientificSoftware/Utility/FCSEXtract/index.htm). Using Mathematica 7 software, we performed logical data transformation and multiple normal distribution component analysis, in order to obtain a more precise estimation of the proportions of different particles. The data transformation in the 'logicle data space' followed published procedures (Parks et al., 2006). Specifically, the scaling function was

\[
S(x;w) = Te^{-(m-w)}(e^{x-w} - p^{2}e^{-(x-w)p} + p^{2} - 1)
\]

where S is the original signal data and x is the corresponding transformed value. T represents the top value of the data, W the width of the linearized values in natural log units, and M the breadth of the display in log decades. Values of parameters were chosen in the following way and applied in the transformation of all the data: The W-value was derived from the largest value of wild type populations (W=0.893747). To allow the display of all wild type values and VGLUT1\textsuperscript{VENUS} values with the same transformation, the R value was set to 1%. The T value was set to the 18-bit maximum value (262,144). The M value used was 4.5. Based on the parameters above, the p-value was 3.69633. The transformed data were binned in steps of 0.02 to allow for an appropriate display of the distributions. Following data transformation, data from 4 independent wild type measurements were fitted as a sum of normal distributions by minimization of the sum of the fit residues (least \(X^{2}\) fit). A good fitting of the data was obtained with a minimum of three normal distributions. The quality of fitting was tested by visualizing the fit residuals, which measure the difference between the fitted function and raw data for each data point. A random distribution of the difference around the x-axis indicates the best fit. Using the resulting 4 wild type probability distribution functions (PDF), an average wild type PDF was computed. The three components of the average wild type PDF were (0.186, 88.7, 25.1), (0.59, 62.4, 11.9), (0.23, 41.8, 11.4), where the values are
proportion, mean, and deviation. To model the data from VGLUT1VENUS synaptosomes, the wild type PDF was incorporated as a fixed part to describe the wild type like, non-fluorescent component in this dataset. Importantly, the proportion of the wild type PDF to other distributions was not fixed. A good fit of the data of VGLUT1VENUS-positive synaptosomes was achieved using the wild type PDF and a minimum of two additional normal distributions. In this way, data from 4 independent experiments were fitted and the mean proportion of each of the 3 components was computed.

**Sample recovery after FASS**

For analysis by Western blotting, electron microscopy, and proteomic experiments, sorted particles were collected and concentrated in a custom built concentrator. This water-cooled device concentrated synaptosomes onto polycarbonate filter discs with a pore size of 0.22 µm by vacuum filtration while keeping the sample at 4°C. During the sort experiment, a minimal volume of 3-5 ml was maintained on top of the filter, to prevent drying of the synaptosomes and allow for efficient cooling in the collection device. After sorting was completed, the walls of the collection device were rinsed once with PBS and subsequently the whole volume was filtered. For biochemical analysis, filters were placed on a pre-cooled glass plate, cut into pieces of roughly 2 mm x 2 mm using a fresh razor blade, transferred to a 2 ml microcentrifuge cup, and flash frozen in liquid nitrogen. Samples were stored at -80°C until further use. For elution of proteins, the appropriate volume of SDS-PAGE sample buffer was added and samples were incubated at 60°C for 20 min under agitation (600 rpm) in a thermomixer. The eluate was then transferred to a fresh 1.5 ml Eppendorf cup.

**Meta-analysis of proteome data**

For proteins that were enriched in or depleted from the VGLUT1VENUS-positive synaptosomes by a factor of two or more, corresponding gene expression data (Cahoy et al, 2008) were downloaded from the NCBI gene expression omnibus (GSE9566) and normalized using the RMA method. To allow retrieval of transcriptome data, proteins identified in the proteomics experiment were matched to their corresponding gene symbols on the Affymetrix 430.2 mouse gene expression array by employing the 'batch-retrieval' function of the Protein Information Resource (http://pir.georgetown.edu/pirwww/search/batch.shtml) (Wu et al, 2003) in combination with the id-converter provided by the Babelomics online tool collection (http://www.babelomics.org) (Al-Shahrour et al, 2008) and NetAffx (http://www.affymetrix.com/analysis/index.affx). Next, an algorithm that clusters genes
according to their pattern of expression in neurons (N), astrocytes (A), and oligodendrocytes (O) was implemented in R (http://www.r-project.org/). For gene clustering, it was first determined if genes are differentially expressed between the cell types studied (Cahoy et al., 2008). A standard deviation of 0.5 or more indicated differential expression. If genes were not differentially expressed, they were divided into three groups according to the level of expression (strong, medium, or weak). The differentially expressed genes were first grouped according to the cell type with maximal mRNA expression. Genes with maximal expression in neurons were grouped in the Nmax group. Consequently the other groups were designated Amax for maximal expression in astrocytes and Omax for maximal expression in oligodendrocytes. Next, the genes within each group were further sub-clustered according to the relation of expression in the remaining two cell types. A standard deviation of 0.7 or more indicated differential expression in the remaining cell types. Finally, genes were ordered according to their mean expression in all cell types. Nmax, Amax, and Omax clusters contained genes with maximal expression in neurons, astrocytes, and oligodendrocytes, respectively, and no difference in expression between the respective remaining cell types; XmaxN, XmaxA, and XmaxO clusters contained genes with maximal expression in cell type X and higher expression in neurons (N), astrocytes (A), or oligodendrocytes (O) than in the remaining cell type. Within each group, genes were ranked according to their average expression level. After clustering of the genes, the Partek genomics suite was used to scale gene expression data, and display relative gene expression levels as a heat map.

The 163 proteins enriched 2-fold or more upon FASS were linked to the corresponding genecards database entry (www.genecards.org; Safran et al., 2010). A few completely unknown proteins were linked to their respective genebank files at NCBI (http://www.ncbi.nlm.nih.gov). Two gene ontology terms for subcellular localization and cellular function were assigned to each protein where possible (The Gene Ontology Consortium, 2000), and proteins were sorted according to function, localization and gene name criteria (Table S4). Distribution bar charts were then generated (Figure 5DE).

**Handling of digital images**

Unless specified otherwise, digital images were processed as follows. Raw digital images were obtained as Tagged Image File Format (TIFF) files from CCD cameras, confocal microscopes, or professional scanners. Linear adjustments of intensity levels were applied uniformly to the whole image when necessary for display purposes. When required for display purposes, regions of interest were cropped and images were converted to 8 bits format.
prior to insertion in figures. Raw data in 16 bits formats were used for quantifications. Image processing was performed with ImageJ (Schneider et al, 2012) and Photoshop CS3 (Adobe Systems Inc, USA) software.

Abbreviations List
A  Astrocytes
CF  Co-Enrichment Factor
Cpx  Complexin
DIV  Day In Vitro
evt/s  Events per Second
FASS  Fluorescence Activated Synaptosome Sorting
FM  FM 4-64
FSC  Forward Scatter
H  Homogenate
LP  Lyzed Pellet
LS  Lyzed Supernatant
MS  Mass Spectrometry
max  Maximum
N  Neurons
NL  Neuroligin
NMDA  N-Methyl-D-Aspartate
O  Oligodendrocytes
P  Pellet
PDF  Probability Distribution Function
PSD  Postsynaptic Density
S  Supernatant
SV  Synaptic Vesicles
WT  Wild Type
Supplemental figure legends

**Fig. S1.** Modeling of flow-data as a combination of normal distributions. The fluorescence intensity distribution of the non-selective sort of all particles and the sort of VGLUT1<sub>Venus</sub>-fluorescent particles was modeled as a mixture of multiple normal distribution components (data transformation in the 'logicle'-display;(Parks et al, 2006)). (A-B) Data from four independent measurements were fitted separately. A minimum of 3 normal distribution components were necessary to produce a best fit. The averaged WT probability distribution function (PDF) represents a WT-like component in the fitting of VGLUT1<sub>Venus</sub> synaptosome data. Two additional normal distributions were necessary to produce an optimal fit for the data from 'all-particles' and the 'VGLUT1<sub>Venus</sub>-fluorescent' populations. The three populations were designated as 'WT-like', 'intermediate', and 'fluorescent', according to their mean fluorescence signals. The fluorescence threshold used in the analysis of flow cytometry data in sorting of VGLUT1<sub>Venus</sub>-positive synaptosomes is indicated. Graphs display a representative fit of the data, and the fit-residual in the scatter plots at the top of each graph. The fit-residual calculates the difference between the actual data and the fitted data to indicate the quality of the fit. A random distribution of the difference around the x-axis indicates that an optimal fit was achieved. (C) The absolute change in the proportion of each of the three populations was calculated as the ratio of 'sorted-fluorescent' over 'sorted-all' particles. For ratios below one, the reciprocal value was plotted in the negative direction to indicate depletion. The change in the relative proportion of the fluorescent component to the other components was also determined.

**Fig. S2.** (A-B) Spectral countings were compared with mRNA expression data of neurons (Neuro), oligodendrocytes (Oligo), and astrocytes (Astro). For enriched and depleted proteins, the corresponding mRNA expression levels in oligodendrocytes, astrocytes, and neurons were compared. Clusters of genes were generated as follows: Xmax clusters contained genes with maximal expression in one cell type over the two others; XmaxY clusters contained genes with a gradation of expression in the three cell types, X or Y being N for neurons, O for oligodendrocytes or A for astrocytes. Note the prevalence of Nmax genes and of glutamatergic markers in the FASS enriched cluster. Genes expressed similarly in all cell types were grouped according to expression level into strong, medium, and weak or not detected (n.d.). Genes whose proteins where analyzed by Western blotting or
immunofluorescence staining in this study are shown in bold. The two candidate proteins Tpd52 and FXYD6 are in bold-red font (see also Tables S1-S3).

**Fig. S3.** Additional immunofluorescence and immunogold electron microscopic analyses of FXYD6 localization. (A) Double immunostaining of primary hippocampal cultures shows the close apposition (arrowheads) of FXYD6 (red) and VGLUT2 (green). Scale bar, 2 µm in overviews and 0.4 µm for enlarged images. (B) Double immunostaining of primary hippocampal cultures show that FXYD6 (red) is not detectable in GFAP-positive (green) astrocytes. Scale bar, 10 µm. (C) Double immunofluorescence staining for FXYD6 (red) and GAD67 (green). In the hippocampus (HPC) stratum radiatum, FXYD6 is detected at the membrane of a GAD67-positive cell. In the cerebellum (cb), FXYD6 is strongly stained in excitatory granule cells, but not in inhibitory Purkinje cells. Scale bar, 5 µm. (D) Data from pre-embedding immunoelectron microscopy analyses of ultrathin sections of mouse striatum using the anti-FXYD6 antibody. Note FXYD6 staining in axons and dendrites. Double arrows point to synaptic contact. St, striatum; HPC, hippocampus; t, terminal; a, axon; d, dendrite. Scale bars, 0.5 µm (1), 1 µm (2), 0.25 µm (3-4).

**Fig. S4.** Immunofluorescence Analysis of Tpd52 Localization in Cultured Hippocampal Neurons. (A-B) Triple immunofluorescence staining for VGLUT1 (green), Tpd52 (red), and either GAD65 (A; blue) or MAP2 (B; blue) in primary hippocampal cultures. Arrowheads show examples of co-localization of Tpd52 and VGLUT1 and lack of co-localization of Tpd52 and GAD65. Scale bars, 1 µm in enlarged images and 6 µm in overviews.
**Table S1.** Mascot results of the proteomic analysis of S-synaptosomes. Identified proteins are sorted alphabetically by protein name. For each protein, column 1 displays the protein name, column 2 displays the gi-number, column 3 displays the Mascot protein hit number, column 4 displays the Mascot protein score, column 5 displays the theoretical molecular weight in kDa, column 6 displays the number of identified “red bold” peptides (i.e. top scoring peptide match for this specific MS/MS spectrum the first time this spectrum has been used in the search result report) and column 7 displays the number of identified non-redundant “red bold” peptides.

**Table S2.** Mascot results of the proteomic analysis of FASS sorted VGLUT1VENUS-positive synaptosomes. Identified proteins are sorted alphabetically by protein name. For each protein, column 1 displays the protein name, column 2 displays the gi-number, column 3 displays the Mascot protein hit number, column 4 displays the Mascot protein score, and column 5 displays the theoretical molecular weight in kDa. Column 6 displays the number of identified “red bold” peptides (i.e. top scoring peptide match for this specific MS/MS spectrum the first time this spectrum has been used in the search result report) and column 7 displays the number of identified non-redundant “red bold” peptides.

**Table S3.** Results of spectral counting using Scaffold. Compilation of the results of Scaffold spectral counting, combined with cell type-specific transcriptome data from Cahoy and collaborators (Cahoy et al, 2008). All proteins quantified by spectral counting are listed, including gi-number, gene symbol, and fold change in spectral counting (FC). Normalized log-scale mRNA expression in neurons, astrocytes, and oligodendrocytes is listed for all proteins that matched with an mRNA identified by Cahoy et al. (2008). For both unsorted and FASS samples, the spectral count (SC), number of peptides (#pept), and protein identification probability (pID) is given. Analysis of selected proteins by Western blotting (WB) or immunofluorescence staining (IF) in the present study is also mentioned.

**Table S4.** Classification of proteins enriched 2-fold or more in VGLUT1VENUS-positive synaptosomes. Protein names were linked to the ad-hoc genecard (www.genecards.org; Safran et al, 2010). A few proteins were linked to their GenBank entry when genecards were not available. Available data were used to assign appropriate GO terms characterizing the subcellular localization and cellular function of proteins. 22 proteins are of unknown function and localization, while others are of known subcellular localization but unknown function.
General GO categories were used to provide a global view on the sample rather than a selective view on individual proteins. The first sheet presents the database of the 163 proteins enriched 2-fold or more in VGLUT1VENUS-positive synaptosomes. The second sheet presents a sorting of the protein names according to function and subcellular localization.

**Supplemental References**


Hebb CO & Whittaker VP (1958) Intracellular distributions of acetylcholine and choline acetylase. *J Physiol (Lond)* **142**: 187–196


