Published in final edited form as:
Nat Neurosci. 2008 July ; 11(7): 799–806. doi:10.1038/nn.2135.

Evolutionary expansion and anatomical specialization of synapse proteome complexity

Richard D. Emes[1,*], **Andrew J. Pocklington**[2,*], **Christopher N.G. Anderson**[3,*], **Alex Bayes**[3], **Mark O. Collins**[3], **Catherine A. Vickers**[4,5], **Mike D.R. Croning**[3], **Bilal R. Malik**[2], **Jyoti S. Choudhary**[3], **J. Douglas Armstrong**[2], and **Seth G.N. Grant**[3]

[1]Institute for Science and Technology in Medicine, Keele University, Thornburrow Drive, Hartshill, Stoke-on-Trent, ST4 7QB, United Kingdom

[2]Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, 5 Forrest Hill, EH1 2QL, United Kingdom

[3]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom

[4]Division of Neuroscience, University of Edinburgh, 1 George Square, Edinburgh, EH8 9JZ, United Kingdom

[5]Okinawa Institute of Science and Technology, Neurobiology Research Unit, Suzaki, Uruma, Okinawa, Japan

Abstract

Understanding the origins and evolution of synapses may provide insight into species diversity and organisation of the brain. Using comparative proteomics and genomics we examined the evolution of the postsynaptic density (PSD) and MAGUK associated signalling complexes (MASCs) underlying learning and memory. PSD/MASC orthologues found in yeast perform basic cellular functions regulating protein synthesis and structural plasticity. Striking changes in signalling complexity were observed at the yeast:metazoan and invertebrate:vertebrate boundaries, with expansion of key synapse components, notably receptors, adhesion/cytoskeletal and scaffold proteins. Proteomic comparison of *Drosophila* and mouse MASCs revealed species-specific adaptation with greater signalling complexity in mouse. Although synapse components were conserved amongst diverse vertebrate species, mapping mRNA and protein expression within the mouse brain showed vertebrate-specific components preferentially contributed to differences between brain regions. We propose that evolution of synapse complexity around a core proto-synapse has contributed to invertebrate–vertebrate differences and to brain specialisation.

Introduction

Learning and adaptation to changing environments are properties shared by all animals and may be involved in adaptive radiation of species into environmental niches. Studies in a wide range of multi-cellular organisms show that simple forms of learning such as sensitisation and habituation found in invertebrates are building blocks for more complex

Corresponding Author:- Dr. Seth G N Grant Wellcome Trust Sanger Institute Hinxton, Cambridge CB10 1SA, United Kingdom Ph: +44 (0) 1223 494 908 Fax: +44 (0) 1223 494 919 sg3@sanger.ac.uk.
*These authors contributed equally

Author Contributions: RDE conducted bioinformatic analysis of MASC/PSD and fMASC. AJP conducted bioinformatic and statistical analyses of MASC/PSD, fMASC and expression datasets. CNGA and CAV conducted RNA and protein expression studies. BRM, AB and MOC conducted fly proteomic studies. MDRC provided informatic support. RDE, AJP, CGNA, JDA and SGNG interpreted the results and prepared the manuscript. SGNG supervised the project.

Europe PMC Funders Author Manuscripts

forms found in vertebrates[1, 2]. Attempts to uncover the neurobiological basis of behavioural complexity have focused on differences between vertebrates, where a high degree of behavioural flexibility has evolved (apparently independently) multiple times (e.g. corvids, cetaceans and primates). Of the factors proposed to explain species differences, it has been argued that those reflecting information-processing capacity (number of cortical neurons and conduction velocity of cortical fibres) correlate best with intelligence amongst mammals[3]. While synapses play a fundamental role in neural information processing, discussions of brain and behavioural evolution typically do not consider the possibility of synapse molecular evolution, of which surprisingly little is known.

The biology of learning and other cognitive functions involves the activation of neurotransmitter receptors on the postsynaptic side of the synapse by patterns of neuronal activity, triggering biochemical pathways leading to changes in neuronal function[1]. The Post Synaptic Density (PSD) contains multi-protein signal transduction complexes formed by neurotransmitter receptors and associated proteins, which are essential for induction of synaptic plasticity and learning[4-7]. Proteomic studies in mice show that ionotropic N-methyl-D-aspartate (NMDA) and metabotropic subtypes of glutamate receptors are linked by scaffold proteins (Membrane Associated Guanylate Kinases, MAGUK) into complexes of 186 proteins referred to as NRC or MASC (NMDA Receptor Complex, MAGUK Associated Signalling Complex)[8-10]. The PSD itself displays a remarkable degree of complexity, involving ~1000 identified proteins from a wide variety of functional classes[11-13]. A high percentage of MASC and PSD genes possess physiological, behavioural and disease phenotypes, with single gene studies showing over 40 MASC proteins to be involved in synaptic and behavioural plasticity in rodents, and human brain disorders ([14-16] http://www.genes2cognition.org/db.html). These proteins and complexes are a suitable template for examining the molecular evolution of learning and synapse organisation.

Here we examine synapse evolution using genomic, proteomic and expression profiling of postsynaptic proteins. Genomic comparison of 19 species indicates differences in the complexity and organisation of synapses. This was confirmed by proteomic studies of Drosophila MASC, which revealed significant differences in complexity compared to mouse MASC. The expression profiles of postsynaptic proteins in mouse brain showed a relationship between the evolution of synapse proteins and their pattern of expression. We present a model for the molecular origins and evolutionary diversification of the synapse, highlighting the roles played by molecular complexity in signal processing and behaviour.

Evolution of the synapse proteome

To investigate the origins of the mammalian synapse proteome, we identified orthologues of 651 genes corresponding to mouse postsynaptic proteins (570 from the PSD,183 from MASC, 102 common to both) (see Methods and[8, 10]) in 19 different species (Supplementary Tables 1 and 2). The species studied comprised a wide range of animals with nervous systems of differing anatomical complexity: invertebrates, non-mammalian vertebrates and mammals. We also identified orthologues in an out-group possessing no nervous system, the unicellular eukaryote *S. cerevisiae*. Validation was performed (repeating orthologue searches by hand using key pairs of organisms and a deeper substitution matrix) to ensure that absence of orthologues did not simply reflect sequence divergence between species (see Supplementary Methods).

The numbers of PSD and MASC orthologues in each species correlate well with each other ($R^2 = 0.99$), and both displayed clear differences between yeast, invertebrates and vertebrates (Figure 1a/b). Approximately 23% of all mammalian synapse proteins were detected in yeast (21.2% MASC, 25.0% PSD) and ~45% in invertebrates (46.2% MASC,

44.8% PSD). Thus a significant proportion of genes encoding MASC and PSD orthologues precede the origins of the nervous system, with apparent stepwise expansions following the divergence of metazoans from eukaryotes and vertebrates from invertebrates. The inclusion of a urochordate (*C. intestinalis*), the closest living relatives of vertebrates[17], allows us to locate this second expansion quite precisely, although analysis of a member of the family Myxinidae (hagfish) would be necessary to distinguish between vertebrate and craniate specific expansion. We would not expect analysis of additional unicellular eukaryotes to significantly alter the picture of metazoan expansion as it primarily involves elaboration of intercellular signalling pathways (fundamental to the existence of metazoa) while genes linked to protein synthesis and metabolism are largely identifiable in yeast (see below).

Synaptic plasticity and learning involves the modulation of cellular function and morphology by receptor-associated signaling pathways. To identify unicellular processes from which synapse functionality may have arisen, we considered the function of PSD/MASC orthologues in yeast as described by the *Saccharomyces* Genome Database (http://www.yeastgenome.org/) (Supplementary Table 3). The majority contribute to generic cellular functions (protein synthesis/degradation, vesicular trafficking, regulation of the actin cytoskeleton) that are regulated in response to environmental factors (ions, nutrients, pheromones). Signal transduction pathways mediating environmental responses contained ~15% of orthologues. Although (predictably) neurotransmitter receptors were absent, other aspects of synaptic signalling possessed functional counterparts in yeast. Orthologues of calmodulin (*cmd1*) and calcineurin (*cmp2*) control Ca^{2+} homeostasis by regulating transcription of *pmc1*[16], orthologue of the synaptic Ca^{2+} pump *ATP2B4*. Independently of Ca^{2+}, *cmd1* regulates the actin cytoskeleton (through interaction with the Arp2/3 complex, also found in the PSD) and receptor-mediated endocytosis[18]. Orthologues of *NF1* (*ira2*), *PKA* (*tpk2*), *Erk2* (*fus3*) and *GNB5* (*ste4*) belong to major pathways regulating transcription, cell morphology and adhesion downstream of nutrient- and pheromone-sensitive GPCRs[19-22]. Thus components of synaptic pathways regulating protein synthesis and structural plasticity in rodents play analogous roles in unicellular responses to environmental cues (ions, nutrients) and simple cell-cell (pheromonal) communication.

In parallel with the expansion in numbers of proteins, the total number of protein domains detected within each set of orthologues increased (Supplementary Figure 1a). However, the number of domain *types* did not increase to the same extent, the difference between invertebrates and vertebrates being much less pronounced (Supplementary Figure 1b). These data suggest that synapse proteome expansion does not so much reflect the recruitment of proteins containing new domain types but more the expansion of protein types already present i.e. innovation by gene family duplication and diversification rather than integration or *de novo* generation of new protein types (see Supplementary Notes and Supplementary Table 4 for further analysis of domain types).

We next examined whether specific types of proteins were involved in synapse proteome expansion. Gene Ontology annotations (http://www.geneontology.org/) were used to evaluate the number of PSD/MASC genes associated with synapse functionality and other more general cell-biological processes in yeast, worm, fly, zebrafish, chicken, mouse and human (Figure 1c). Neurotransmitter receptors were present in all organisms with nervous systems, and vertebrates consistently displayed greater numbers of receptors than invertebrates. The components of second messenger pathways also showed expansion in vertebrates and invertebrates. In contrast, the representation of protein synthesis machinery did not increase. In agreement with the GO term analysis, functional families corresponding to upstream signalling/structural components were poorly represented in yeast, undergoing increasing expansion in invertebrates and vertebrates (Figure 1d). As predicted, the majority of components from downstream cell biological processes could be identified in yeast and

early metazoans, and showed significantly reduced expansion (Figure 1e). Of the upstream components, 44% of all cytoskeletal and cell adhesion molecules were vertebrate in origin, significantly more than expected from a random sample of genes ($P < 10^{-4}$, see Methods). Only 10% of cytoskeletal/adhesion genes possessed detectable orthologues in yeast ($P < 10^{-4}$). Also under-represented in yeast (10% or less identified) were channels and receptors ($P = 0.003$), MAGUKs/Adaptors/Scaffolders ($P = 0.02$) and kinases ($P = 0.05$). Downstream processes comprised the majority of synaptic components with identifiable orthologues in yeast. Notably enriched were genes linked to transcription and translation (over 50%, $P < 10^{-3}$), including 85% of ribosomal proteins ($P < 10^{-5}$); protein folding and trafficking (70% of heat shock/chaperones, $P < 10^{-5}$); and metabolism (90% of ATP synthases, $P < 10^{-4}$). So while signal transduction pathways linked to cell surface receptors show evidence of expansion at yeast:metazoan and invertebrate:vertebrate boundaries, the downstream cell-biological processes that they regulate do not.

These data suggest that most functional types of synapse protein were present in early metazoans, and that the proto-synapse constructed from this core functionality has been elaborated upon during the evolution of invertebrates and vertebrates. Recent studies confirm the presence of many families of synapse and cell signalling genes in the phylum porifera (sponges) supporting the hypothesis that core synapse signalling components were present at the base of animal kingdom[23-25]. Elaboration appears to have primarily involved gene family expansion and diversification amongst upstream signalling/structural components (receptors, scaffolders, cytoskeletal, adhesion and signal transduction molecules).

Key predictions arising from these comparative genomic data are that invertebrate synapses a) exhibit reduced signalling complexity (number of signalling/structural components) compared to vertebrates b) possess components of generic cell-biological processes (e.g. protein synthesis, metabolism) that are predominantly of pre-metazoan origin, and c) show evolutionary expansion in upstream signalling/structural components. To test these predictions it is necessary to have proteomic data from invertebrate synapses and although PSDs were observed with electron microscopy in the fly brain[26] they have not been analysed at the molecular level. However, the *discs large* protein (Dlg), which is a Drosophila MAGUK and the homologue of mammalian PSD-95, SAP102 and PSD-93 has been studied and is expressed widely in the fly nervous system[27]. Therefore we isolated Drosophila MASC complexes, translating to fly the methods used to isolate mouse MASC.

Isolation of Drosophila MAGUK Associated Signalling Complexes

In line with the method for purifying mouse MASC (mMASC)[28], we generated a C-terminus hexapeptide of the *D. melanogaster* NR2 subunit (dPEP6; EMETVL), and a control hexapeptide lacking the PDZ interaction motif (dPEP6ΔVL; IAEMET see supplementary methods), for affinity purification of protein complexes. We found that affinity columns using dPEP6, but not dPEP6ΔVL, bound Dlg as shown using immunoblotting (Figure 2a). Moreover, coomassie stained gels showed that dPEP6 columns retrieved many proteins, and that the DLG2 band was notably absent from dPEP6ΔVL columns (Figure 2b). We therefore concluded that dPEP6 was capable of isolating fly MAGUK proteins and MASCs (fMASC). 220 fMASC proteins were identified using mass spectrometry (see Methods and Supplementary Table 5), suggesting that *Drosophila* may possess complexes of a size comparable to those in mouse[8, 10]. At first glance this appeared to contradict the comparative genomic data, but closer inspection revealed major differences in the types of proteins present. When fMASC proteins were categorized into functional protein families using the scheme developed for mMASC[8], it was found that upstream signalling/structural components (receptors, scaffolders, signal transduction molecules, etc.)

accounted for ~25% of fMASC proteins, compared to >60% of mMASC (see Figure 2c, note that fMASC 'Signalling molecules and Enzymes' are predominantly metabolic enzymes, heat shock/chaperones and mitochondrial proteins). Thus the molecular complexity of MASC *signalling* in fly is roughly half that in mouse (in both relative and absolute terms), as predicted.

To test the remaining predictions, we next asked which types of fMASC components (if any) showed evidence of evolutionary conservation/expansion when compared to yeast. We also noted whether expansion preceded the divergence of fly and mouse lineages (the corresponding genes being detectable in chordates), or whether it was fly-specific. Orthologues of fMASC genes were identified in yeast and in the chordate species previously analysed (Supplementary Table 6). 71% of fMASC genes were identified in yeast, only 64 (29%) appearing to be of metazoan origin. In agreement with our predictions, the vast majority of downstream components were present in yeast (Figure 2e, 'Yeast' column), while upstream signalling/structural components of fMASC showed fly-specific expansion (Figure 2d, 'Fly-specific' column). Thus both fMASC and mMASC appear to have undergone lineage specific adaptation.

We were interested in the extent to which fMASC and mMASC represent similar synaptic sub-components by comparing the composition of fMASC to mMASC and mPSD. We focused on proteins with a primary role in synaptic function/signalling as opposed to protein metabolism/synthesis or other cellular processes (e.g. mitochondrial function) (Supplementary Tables 5 and 7). 67 (30%) fMASC proteins were identified as having primarily synaptic function, in contrast with 155 (83%) mMASC and 355 (62%) mPSD (85 of which were also present in mMASC, see Supplementary Table 7). fMASC and mMASC appeared to contain similar proportions of each functional family of proteins (suggesting a similar type of signalling complex), while fMASC and mPSD showed less similarity. To quantify this, we calculated the probability of a random set of 67 proteins from mMASC or mPSD having the same functional representation (number of proteins in each functional class) as fMASC. The probability of obtaining the fMASC representation from mMASC was 80 times greater than the probability of obtaining it from mPSD, and over 8000 times greater than obtaining it from the subset of mPSD not found in mMASC. Thus in terms of their general composition, fMASC and mMASC comprise similar functional sub-components of the synapse.

Finally we considered the genetic similarity of fMASC and mMASC, and whether they showed evidence of having evolved from a common synaptic sub-component. Given the extent of both fly-specific and vertebrate-specific adaptation, we would not expect to find a high degree of molecular identity between fMASC and mMASC. However, we would expect to find greater evidence of expansion from common ancestral genes amongst synaptic signalling molecules. 44 (20%) fMASC components were identified as being orthologous to 56 mMASC/mPSD genes. Of these, 40 (71%) displayed evidence of gene family expansion (1-many, many-many gene mapping) between fly and mouse, including 85% of all fMASC synaptic function gene orthologues ($P = 0.01$). Using BLAST[29] to identify more distant relationships, we found that 62 (93%) fMASC synaptic function genes had identifiable homologues in the mouse genome, with 50 (75%) possessing synaptic function homologues in mMASC and/or mPSD, (see Supplementary Methods). Supporting the identification of fMASC and mMASC as similar synaptic sub-components, 33 of the 50 were present in mMASC. Thus in terms of their isolation, functional composition and phylogenetic relationship, fMASC and mMASC appear to represent species-specific adaptations of a common synaptic sub-component, having diverged by duplication, recruitment and replacement of genes.

It seems highly plausible that the dramatic increase in molecular signalling complexity noted in vertebrates contributes to their increased capacity for behavioural complexity. Even relatively small changes in the number of different signalling components may have a large multiplicative impact on neuronal function. For example, as invertebrates have single NR2 and Dlg genes whereas vertebrates have 4 of each, the number of potential NR2-Dlg complexes in vertebrates is 16 times higher. Extending this to other expanded families of synaptic proteins suggests a major increase in the number of combinations with distinct functional properties that are available to the vertebrate nervous system.

While molecular complexity clearly contributes to neuronal function, it does so within the context of an expanded mammalian nervous system that has undergone significant regional specialisation. Indeed the genome duplication events leading to more complex synapse proteomes in vertebrates predate the origins of species with anatomically large nervous systems (all vertebrate species had identifiable orthologues of ~80% or more mouse MASC/PSD genes, Figure 1, 4). This immediately leads to the question as to whether the complexity of the synapse proteome has been utilised in the expansion and diversification of vertebrate nervous systems, and whether anatomical regionalisation has been accompanied by synapse specialisation through the tuning of expression of synapse proteome components.

MASC expression diversity in mouse brain

We therefore examined the expression patterns of synaptic proteins in different regions of the mouse brain. In total we examined >150 molecules and 22 brain regions using protein and mRNA assays. Protein: western blotting (WB) of tissue extracts probed with antibodies to 65 different synaptic proteins, 56 of which are present in the MASC; immunohistochemistry (IHC) of mouse brain sections with antibodies to 43 different synaptic proteins, 39 of which were MASC proteins. mRNA: in situ hybridization (ISH) results for 55 MASC genes were obtained from the Brain Gene Expression Map (BGEM, http://www.stjudebgem.org,[30]; microarray (MA) data for 148 different MASC genes (published in [30]). WB, IHC and ISH data was collated for 4 brain regions (hippocampus, cortex, striatum and cerebellum), while MA data was available for a larger set of 22 regions[31]. To facilitate comparison between datasets in the 4 common regions, we used a similar scoring method where each gene in each brain region was assigned an expression score in the range of 0 to 4 (see Supplementary Methods, and Notes for correlation values between the different datasets). Expression data is contained in Supplementary Table 8, while representative IHC images are presented in Supplementary Figure 2.

We found that the vast majority of proteins were clearly co-expressed in each area of the brain investigated, with >95% of genes co-expressed in forebrain structures (hippocampus, striatum and cortex) and >80% expressed in all four regions. However, it was apparent by visual inspection of all datasets that there was variation between the expression patterns of these genes (Supplementary Figure 2, Supplementary Table 8). We therefore classified expression patterns by Anatomical Variation of expression (AVex) of proteins/mRNAs using 4 categories: $AVex^{zero}$, same expression score in each region and therefore zero variation; $AVex^{low}$, difference between highest and lowest expression scores of ≤1 (but not zero); $AVex^{high}$, difference between highest and lowest scores ≥3; and $AVex^{med}$, expression variability lying between $AVex^{high}$ and $AVex^{low}$. It was clear that the majority of genes/proteins were in the variable categories, and for MA, IHC and ISH, the majority of genes/proteins were in the two most variable AVex classes (Figure 3a). These data show that each brain region expressed a similar set of postsynaptic proteins although in different combinations of levels. These profiles or barcodes can be used to identify a particular region and is consistent with previous studies [31-33]. These region-specific expression profiles

indicate that variation in expression of synapse proteome components confers regional specialisation with differential signal processing.

Since proteomic and phylogenetic analyses show that adaptation of synaptic function *between* species primarily involves the tuning of upstream signal processing pathways we investigated the relative contribution of pre-metazoan, invertebrate and vertebrate innovations to regional specialisation of synaptic function, and asked whether specific classes of molecules were involved. Using the microarray data covering expression of 148 MASC genes over 22 brain regions, we calculated the standard deviation in each gene's expression and identified those with high or low variability (see Supplementary Methods). We then plotted the percentage of high/low variability genes found in each of the phylogenetic groupings (Figure 3b). Invertebrate and vertebrate innovations contributed fewer genes to the least variable category, whereas vertebrate innovations contributed greatly to the most variable category. 52% of all highly varying genes were of vertebrate origin, significantly more than expected from a random sample of genes ($P = 0.003$), and 48% of all least variable genes were of pre-metazoan origin ($P = 0.006$). Similar results were observed in the WB, IHC and ISH datasets (Supplementary Figure 3).

These data clearly indicate that genes contributing most to anatomical variation in expression in mouse brain are typically of more recent origin. It was also found that highly varying genes were significantly enriched with upstream signalling/structural components, reflecting their greater expansion in invertebrates and vertebrates. Of the 10 MAGUKs/Adaptors/Scaffolders for which western blot expression was available, 8 were classified as highly varying, accounting for 42% of all $AVex^{high}$ proteins in this dataset ($P = 0.002$), and including all 6 PDZ-domain containing scaffolders ($P = 0.0008$). Similarly, all 7 glutamate receptors studied by IHC were classified as $AVex^{high}$ (41% of all highly varying IHC genes, $P = 0.0007$). In contrast, no members of the Signalling Molecules and Enzymes class (60% of which consisted of ATP synthases, mitochondrial and other enzymes) were found to be highly varying ($P = 0.001$). Within the telencephalon, this class of molecules accounted for 42% of all molecules with least varying microarray expression ($P = 0.006$).

Discussion

Using proteomic, genomic, and expression profiling tools we present a study of the evolution of brain synapses. The data are consistent with a model in which core components of the synapse originated in unicellular eukaryotes where they play roles in responses to environmental stress (Figure 4). Stepwise expansions in molecular signalling complexity coincided with the divergence of metazoans from eukaryotes and vertebrates from invertebrates. Most functional types of synapse protein were present in early metazoans, and elaboration upon this core functionality primarily involved gene family expansion and diversification amongst upstream signalling and structural components (receptors, scaffolders, cytoskeletal, adhesion and signal transduction molecules). Proteomic comparison of vertebrate (mouse) and invertebrate (fly) MASC showed them to be species-specific adaptations of a common synaptic sub-component, having diverged by duplication, recruitment and replacement of genes. Thus significant expansions in complexity of the synapse proteome and specialisation have occurred during diversification.

Expansion of the synapse proteome predates the origin of vertebrate species that have anatomically large nervous systems. The expression patterns of synaptic proteins in mouse brain showed each region expressed a similar set of postsynaptic proteins although in different combinations of levels (Figure 4b). Genes contributing most to anatomical variation in expression were typically of more recent origin, and were significantly enriched with upstream signalling/structural components (reflecting their greater expansion in

invertebrates and vertebrates). This preferential contribution to anatomical diversification in the mammalian brain made by synapse proteome innovations of vertebrate origin suggests that tuning of expression of synapse proteome genes resulted in functional diversification in brain regions.

Evolution of synaptic signalling complexity

How might the increased complexity of postsynaptic signalling complexes in vertebrates compared with invertebrates influence plasticity and behaviour? First, the expansion of upstream proteins including receptors provides a wider range of specificity elements for ligands and extracellular signals. Second, the combinations and organisation of neurotransmitter receptor complexes show significant differences. In particular the NMDA receptor of invertebrates has a single NR2 subunit and single Dlg MAGUK adaptor, thereby forming a single complex. In contrast, vertebrates have 4 of each thus allowing 16 times the number of potential NR2-Dlg complexes. Mutations in mice support the model where the increased molecular complexity of mammalian MASC contributes to diversity in behavioural and electrophysiological signalling. Comparison of Dlg2 (PSD-93), Dlg3 (SAP102) and Dlg4 (PSD-95) mutants show unique synaptic plasticity phenotypes in response to different patterns of neural stimuli in the CA3-CA1 synapses of the hippocampus. At the behavioural level there were distinct cognitive phenotypes in learning tasks for these mutants[4, 6]. Similarly, the distinct mammalian NR2 subunits have specific signalling and behavioural phenotypes[7, 34, 35] as do many (>40) other MASC proteins, drawn from various functional protein classes ([16], http://www.genes2cognition.org/db/).

Complexity differences between invertebrate and vertebrate MASC are accompanied by differences in interaction domains that are responsible for organising the signalling complexes[36]. The vertebrate NR2 cytoplasmic domain, which is responsible for binding MAGUK proteins (via the terminal PDZ interaction motif) and multiple postsynaptic signaling molecules, is ~600 residues[37-40]. In contrast, the invertebrate NR2 subunit possesses a short ~100 residue domain lacking most of these interaction motifs whilst retaining the PDZ motif[36]. These differences in the organisation and complexity of signalling complexes would be expected to result in postsynaptic signalling networks with distinct computational capabilities between vertebrates and invertebrates[16].

Anatomical diversification and evolution of synapse proteome complexity

Comparison of phylogeny and brain expression revealed unexpected relationships – recently evolved genes encoding upstream signalling/structural components of pathways contribute most to anatomical diversity (Figure 4b). Electron microscopy studies of single synapses in rodents reveal that this diversity (for NR2 and MAGUK proteins) distinguishes individual synapses[41, 42]. Although the full extent of synapse diversity is unknown, we observed that the differential anatomical expression patterns of proteins within MASCs produces a signature or expression ‘barcode’. Given the large number of postsynaptic components and their variation in levels there are very large numbers of potential combinations of complexes or synapses. The variation in levels of expression in brain regions is likely to have been driven by mutations within cis-regulatory sequences of expanded vertebrate gene families; a process known as subfunctionalization. Both the comparison of control elements in these synaptic genes and comparative proteomics between brain regions in different species should provide further insight into the mechanisms generating synapse diversity.

Our model of synapse evolution indicates an ancestral or prototype synapse which has been elaborated upon to provide species with a mechanism for anatomical diversification and specialisation of synapse function. It follows that synapse proteome complexity contributes to the computational and cognitive properties of the brain, and should be considered along

with differences in neuron number and connectivity in interpreting evolutionary differences in behaviour. By providing a substrate for synapse diversification, it is interesting to speculate that synapse molecular complexity may have been a prerequisite for anatomical and functional changes underlying the emergence of complex behavioural repertoires. Although the complexity and diversity of human synapse proteomes remains to be explored it is interesting that both ancient (e.g. NF1) and recent components (e.g. Dlg3) of MASC are encoded by genes responsible for heritable cognitive impairments. Moreover, some MASC components (e.g. NR2A) show evidence of higher rates of evolution in primate compared to rodent lineages[43]. The synapse proteome datasets from invertebrates and vertebrates provide a new approach for studying nervous system evolution and diversification and may be used to shed light on the origins of complex behaviours.

Materials and methods

Isolation of Drosophila MASC

Affinity resin preparation—0.5 ml activated support Affi-Gel 10 (Bio-Rad, USA) was washed with 20ml chilled MQ-Water and 5ml of chilled MOPS pH 8.0 50mM $CaCl_2$ 50mM. 5 mg of peptide in 1ml MOPS pH 8.0 50mM $CaCl_2$ 50mM was mixed with Affi-Gel and coupled for 4 hours at 4 °C with a further blocking step (Tris pH9 1M 2:1 (v/v) 18hrs 4°C, washed with Tris 20mM pH9.0 and stored at 4°C.

Affinity purification—1 g *D. melanogaster* heads were homogenized in 24 ml lysis buffer: Tris 50 mM pH 7.4, Nonidet P-40 (Roche Diagnostics, Germany) 0.5% (v/v), NaF 50 mM, $ZnCl_2$ 20 μM, o-vanadate 1 mM, PMSF 1mM, Aprotinin 2ugr/ml and Leupeptin 2 μg/ml, using an OMNI 2000 homogenizer (Omni International, USA) on ice. Sample was left in ice for 1 hour and centrifuged using a JA25.50 rotor (Beckman, USA) at 20,000 rpm for 30 minutes at 4°C and supernatant cleared through a 5 μM filter. Extracts were mixed with affinity resins at a 100:1 (v/v) and incubated 18hrs at 4°C with agitation. Resin was washed with 300 column volumes of lysis buffer lacking PMSF. Resin was mixed with an equivalent volume of elution buffer (5 mg/ml solution of the same peptide contained on the resin in washing buffer with pH adjusted to 9) and incubated for 2 hours at 4°C with agitation. The mixture of resin and elution buffer was centrifuged for 1 minute at 3000rpm in an Eppendorf table centrifuge and the supernatant containing the eluted sample recovered.

Detection of orthologues across species—Swiss-Prot and TrEMBL accessions of PSD and NRC/MASC proteins previously identified by mass spectrometry and immunoprecipitation[8] were mapped to human Ensembl genes from NCBI Build 35 using EnsMart[44]. Orthologues of these were identified in 19 species using the Ensembl Compara database (www.ensembl.org, Ensembl version 36). A full list of gene builds used is given in Supplementary Methods. Pfam domains[45] and Gene Ontology (GO) terms (www.geneontology.org) for each orthologue were obtained via EnsMart. To infer species phylogeny, protein sequences of GAPDH were aligned using muscle[46]. The unrooted maximum likelihood tree was generated by PhyML[47].

Identification of fly MASC orthologues—Orthologues of the 220 fMASC genes together with their orthology type were retrieved from the Ensembl 46 database via EnsMart. In order to identify fMASC genes preceding the divergence of fly and mouse lineages, orthologues were identified in yeast and the 14 chordate species. A full list of gene builds used is given in Supplementary Methods.

Statistical Analysis

Composition of fly MASC relative to mouse MASC/PSD (synaptic function group)—To assess similarity in composition between the synaptic function subsets of fMASC (67 proteins) and mMASC/mPSD (155 and 355 respectively), we calculated the probability of a random set of 67 proteins from mMASC or mPSD having the same number from each functional family as fMASC. Consider a set of N proteins of which n(i) belong to functional family f(i) (i = 1…n_f). We wish to calculate the probability that a random selection of M proteins will contain m(i) in functional family f(i) (i = 1…n_f). There are $\rho[M] = N!/[M!(N-M)!]$ ways of selecting M proteins from N. Of these, $\rho[m(1), m(2), \ldots, m(n_f)] = \Pi_i\, n(i)!/[m(i)![n(i)-m(i)]]!$ contain m(i) in functional family f(i) (i = 1…n_f). The probability is thus $\rho[m(1), m(2), \ldots, m(n_f)]/\rho[M]$.

Statistical overlap between sets of molecules—The statistical significance of an overlap between two sets of molecules was calculated using the method of[16]. This was used to analyse several datasets: phylogenetic expansion of mouse MASC/PSD functional classes; phylogenetic expansion of mouse fMASC functional classes; relationship between functional classes of mMASC/mPSD orthologues of fMASC components and orthology type; expression variability and phylogeny; and expression variability and mouse MASC/PSD functional classes. Further details are given in Supplementary Methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Wellcome Trust for support to CNGA, JDA, RDE, MOC, MDRC, JSC, SGNG; MRC for support to AJP, RDE; GSK for support to BRM, Edinburgh University and e-Science Institute to JDA; EMBO for AB. L. Hansen and L. Valor for assistance with array data; M. Marshall, K. Page for literature curation; members of Genes to Cognition program for discussions and J.V. Turner for editorial assistance.

References

1. Kandel ER. The molecular biology of memory storage: a dialogue between genes and synapses. Science. 2001; 294:1030–8. [PubMed: 11691980]
2. Moore BR. The evolution of learning. Biol Rev Camb Philos Soc. 2004; 79:301–35. [PubMed: 15191226]
3. Roth G, Dicke U. Evolution of the brain and intelligence. Trends Cogn Sci. 2005; 9:250–7. [PubMed: 15866152]
4. Cuthbert PC, et al. Synapse-associated protein 102/dlgh3 couples the NMDA receptor to specific plasticity pathways and learning strategies. J Neurosci. 2007; 27:2673–82. [PubMed: 17344405]
5. Komiyama NH, et al. SynGAP regulates ERK/MAPK signaling, synaptic plasticity, and learning in the complex with postsynaptic density 95 and NMDA receptor. J Neurosci. 2002; 22:9721–32. [PubMed: 12427827]
6. Migaud M, et al. Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein. Nature. 1998; 396:433–9. [PubMed: 9853749]
7. Sprengel R, et al. Importance of the intracellular domain of NR2 subunits for NMDA receptor function in vivo. Cell. 1998; 92:279–89. [PubMed: 9458051]
8. Collins MO, et al. Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. J Neurochem. 2006; 97(Suppl 1):16–23. [PubMed: 16635246]
9. Dosemeci A, et al. Composition of the synaptic PSD-95 complex. Mol Cell Proteomics. 2007

10. Husi H, Ward MA, Choudhary JS, Blackstock WP, Grant SG. Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. Nat Neurosci. 2000; 3:661–9. [PubMed: 10862698]
11. Collins MO, et al. Proteomic analysis of in vivo phosphorylated synaptic proteins. J Biol Chem. 2005; 280:5972–82. [PubMed: 15572359]
12. Yamauchi T. Molecular constituents and phosphorylation-dependent regulation of the post-synaptic density. Mass Spectrom Rev. 2002; 21:266–86. [PubMed: 12533800]
13. Yoshimura Y, et al. Molecular constituents of the postsynaptic density fraction revealed by proteomic analysis using multidimensional liquid chromatography-tandem mass spectrometry. J Neurochem. 2004; 88:759–68. [PubMed: 14720225]
14. Grant SG, Marshall MC, Page KL, Cumiskey MA, Armstrong JD. Synapse proteomics of multiprotein complexes: en route from genes to nervous system diseases. Hum Mol Genet. 2005; 14(Spec No. 2):R225–34. [PubMed: 16150739]
15. Laumonnier F, Cuthbert PC, Grant SG. The role of neuronal complexes in human X-linked brain diseases. Am J Hum Genet. 2007; 80:205–20. [PubMed: 17236127]
16. Pocklington AJ, Cumiskey M, Armstrong JD, Grant SGN. The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour. Mol Syst Biol. 2006; 2:E1–E14.
17. Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature. 2006; 439:965–8. [PubMed: 16495997]
18. Cyert MS. Genetic analysis of calmodulin and its targets in Saccharomyces cerevisiae. Annu Rev Genet. 2001; 35:647–72. [PubMed: 11700296]
19. Elion EA, Qi M, Chen W. Signal transduction. Signaling specificity in yeast. Science. 2005; 307:687–8. [PubMed: 15692041]
20. Erdman S, Snyder M. A filamentous growth response mediated by the yeast mating pathway. Genetics. 2001; 159:919–28. [PubMed: 11729141]
21. Harashima T, Anderson S, Yates JR 3rd, Heitman J. The kelch proteins Gpb1 and Gpb2 inhibit Ras activity via association with the yeast RasGAP neurofibromin homologs Ira1 and Ira2. Mol Cell. 2006; 22:819–30. [PubMed: 16793550]
22. Palecek SP, Parikh AS, Kron SJ. Sensing, signalling and integrating physical processes during Saccharomyces cerevisiae invasive and filamentous growth. Microbiology. 2002; 148:893–907. [PubMed: 11932437]
23. Nichols SA, Dirks W, Pearse JS, King N. Early evolution of animal cell signaling and adhesion genes. Proc Natl Acad Sci U S A. 2006; 103:12451–6. [PubMed: 16891419]
24. Sakarya O, et al. A post-synaptic scaffold at the origin of the animal kingdom. PLoS ONE. 2007; 2:e506. [PubMed: 17551586]
25. te Velthuis AJ, Admiraal JF, Bagowski CP. Molecular evolution of the MAGUK family in metazoan genomes. BMC Evol Biol. 2007; 7:129. [PubMed: 17678554]
26. Yasuyama K, Meinertzhagen IA, Schurmann FW. Synaptic organization of the mushroom body calyx in Drosophila melanogaster. J Comp Neurol. 2002; 445:211–26. [PubMed: 11920702]
27. Ruiz-Canada C, Koh YH, Budnik V, Tejedor FJ. DLG differentially localizes Shaker K+-channels in the central nervous system and retina of Drosophila. J Neurochem. 2002; 82:1490–501. [PubMed: 12354297]
28. Husi H, Grant SG. Isolation of 2000-kDa complexes of N-methyl-D-aspartate receptor and postsynaptic density 95 from mouse brain. J Neurochem. 2001; 77:281–91. [PubMed: 11279284]
29. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–402. [PubMed: 9254694]
30. Magdaleno S, et al. BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system. PLoS Biol. 2006; 4:e86. [PubMed: 16602821]
31. Zapala MA, et al. Adult mouse brain gene expression patterns bear an embryologic imprint. Proc Natl Acad Sci U S A. 2005; 102:10357–62. [PubMed: 16002470]

32. Lein ES, Zhao X, Gage FH. Defining a molecular atlas of the hippocampus using DNA microarrays and high-throughput in situ hybridization. J Neurosci. 2004; 24:3879–89. [PubMed: 15084669]
33. Toledo-Rodriguez M, et al. Correlation maps allow neuronal electrical properties to be predicted from single-cell gene expression profiles in rat neocortex. Cereb Cortex. 2004; 14:1310–27. [PubMed: 15192011]
34. Kutsuwada T, et al. Impairment of suckling response, trigeminal neuronal pattern formation, and hippocampal LTD in NMDA receptor epsilon 2 subunit mutant mice. Neuron. 1996; 16:333–44. [PubMed: 8789948]
35. Sakimura K, et al. Reduced hippocampal LTP and spatial learning in mice lacking NMDA receptor epsilon 1 subunit. Nature. 1995; 373:151–5. [PubMed: 7816096]
36. Ryan TJ, Emes RD, Grant SG, Komiyama NH. Evolution of NMDA receptor cytoplasmic interaction domains: implications for organisation of synaptic signalling complexes. BMC Neurosci. 2008; 9:6. [PubMed: 18197970]
37. Chung HJ, Huang YH, Lau LF, Huganir RL. Regulation of the NMDA receptor complex and trafficking by activity-dependent phosphorylation of the NR2B subunit PDZ ligand. J Neurosci. 2004; 24:10248–59. [PubMed: 15537897]
38. Li BS, et al. Regulation of NMDA receptors by cyclin-dependent kinase-5. Proc Natl Acad Sci U S A. 2001; 98:12742–7. [PubMed: 11675505]
39. Nakazawa T, et al. NR2B tyrosine phosphorylation modulates fear learning as well as amygdaloid synaptic plasticity. Embo J. 2006; 25:2867–77. [PubMed: 16710293]
40. Prybylowski K, et al. The synaptic localization of NR2B-containing NMDA receptors is controlled by interactions with PDZ proteins and AP-2. Neuron. 2005; 47:845–57. [PubMed: 16157279]
41. Petralia RS, Sans N, Wang YX, Wenthold RJ. Ontogeny of postsynaptic density proteins at glutamatergic synapses. Mol Cell Neurosci. 2005; 29:436–52. [PubMed: 15894489]
42. Sans N, et al. A developmental change in NMDA receptor-associated proteins at hippocampal synapses. J Neurosci. 2000; 20:1260–71. [PubMed: 10648730]
43. Dorus S, et al. Accelerated evolution of nervous system genes in the origin of Homo sapiens. Cell. 2004; 119:1027–40. [PubMed: 15620360]
44. Kasprzyk A, et al. EnsMart: a generic system for fast and flexible access to biological data. Genome Res. 2004; 14:160–9. [PubMed: 14707178]
45. Bateman A, et al. The Pfam protein families database. Nucleic Acids Res. 2000; 28:263–6. [PubMed: 10592242]
46. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–7. [PubMed: 15034147]
47. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52:696–704. [PubMed: 14530136]
48. White JG, Southgate E, Thomson JN, Brenner S. The structure of the nervous system of the nematode C. elegans. Philos. Trans. Royal Soc. Lond. B Biol. Sci. 1986; 314:1–340.
49. Truman, JW.; Taylor, BJ.; Awad, TA. The Development of Drosophila melanogaster. Bate, M.; Martinez Arias, A., editors. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 1993. p. 1245-1276.
50. Imai JH, Meinertzhagen IA. Neurons of the ascidian larval nervous system in Ciona intestinalis: I. Central nervous system. J Comp Neurol. 2007; 501:316–34. [PubMed: 17245701]

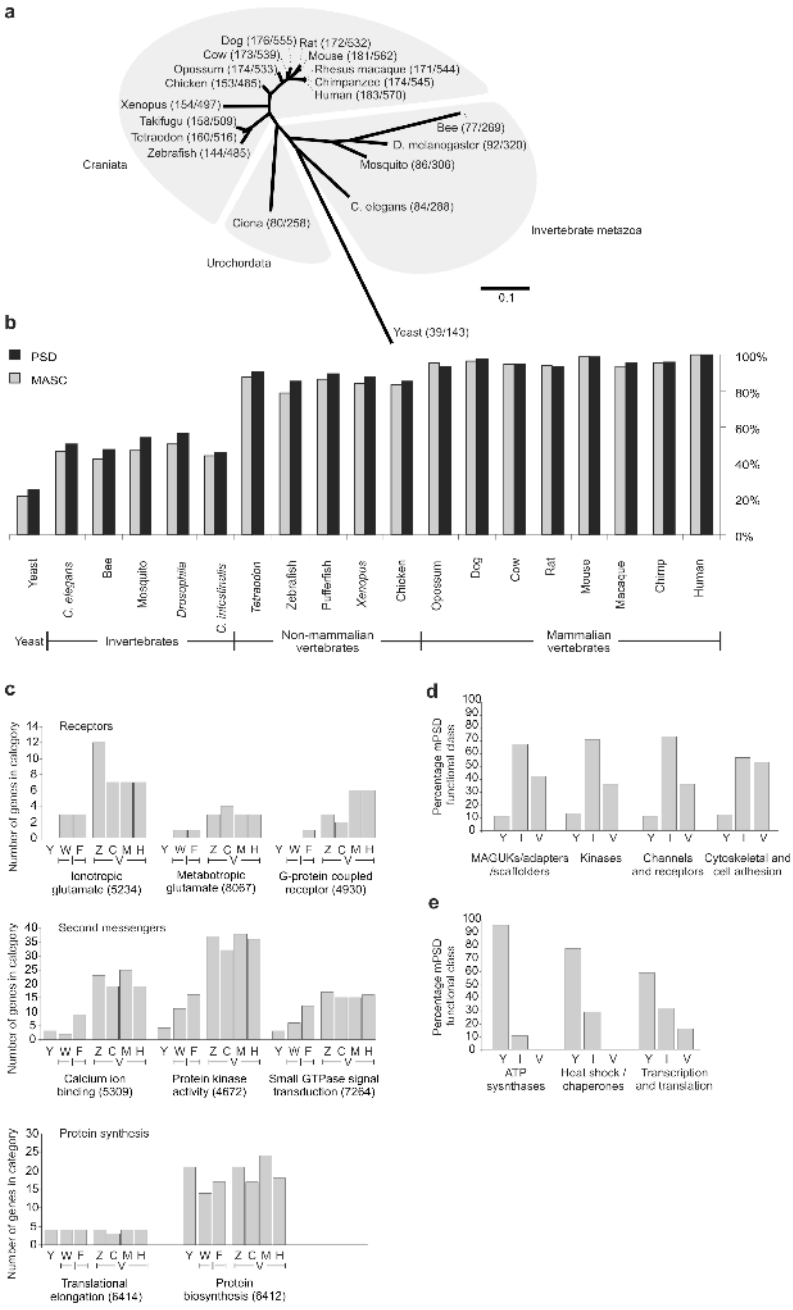

Figure 1. Comparison of PSD and MASC homologues and expansion of selected functional groups of genes

a) The phylogenetic relationship of species studied. Numbers in parentheses represent number of NRC/PSD orthologues detected respectively.

b) The occurrence of PSD and MASC homologues found in each of the 19 species as a percentage of those found in human. Where annotation of multiple homologues was reported by Ensembl, a single positive hit was recorded.

c) Evolution of learning and plasticity mechanisms. From the 651 PSD/MASC genes, the number of genes in different species (yeast (Y), worm (W), fly (F), zebrafish (Z), chicken (C), mouse (M) and human (H)) involved with 3 major molecular mechanisms (receptors,

second messenger signaling, protein synthesis) of learning and memory were plotted. For clarity, only data from seven representative species are shown. Numbers in parentheses are GO term identifiers (www.geneontology.org/). Data were obtained from Ensembl.
d) Upstream signaling components show increasing rates of expansion towards mammalian lineage. The proportion of each functional class (as a percentage of the total number of mouse MASC/PSD genes belonging to it) whose earliest identifiable orthologue occurs in yeast (Y), invertebrates (I) or vertebrates (V).
e) Downstream signaling components show decreasing rates of expansion towards mammalian lineage. The proportion of each functional class (as a percentage of the total number of mouse MASC/PSD genes belonging to it) whose earliest identifiable orthologue occurs in yeast (Y), invertebrates (I) or vertebrates (V).

Europe PMC Funders Author Manuscripts

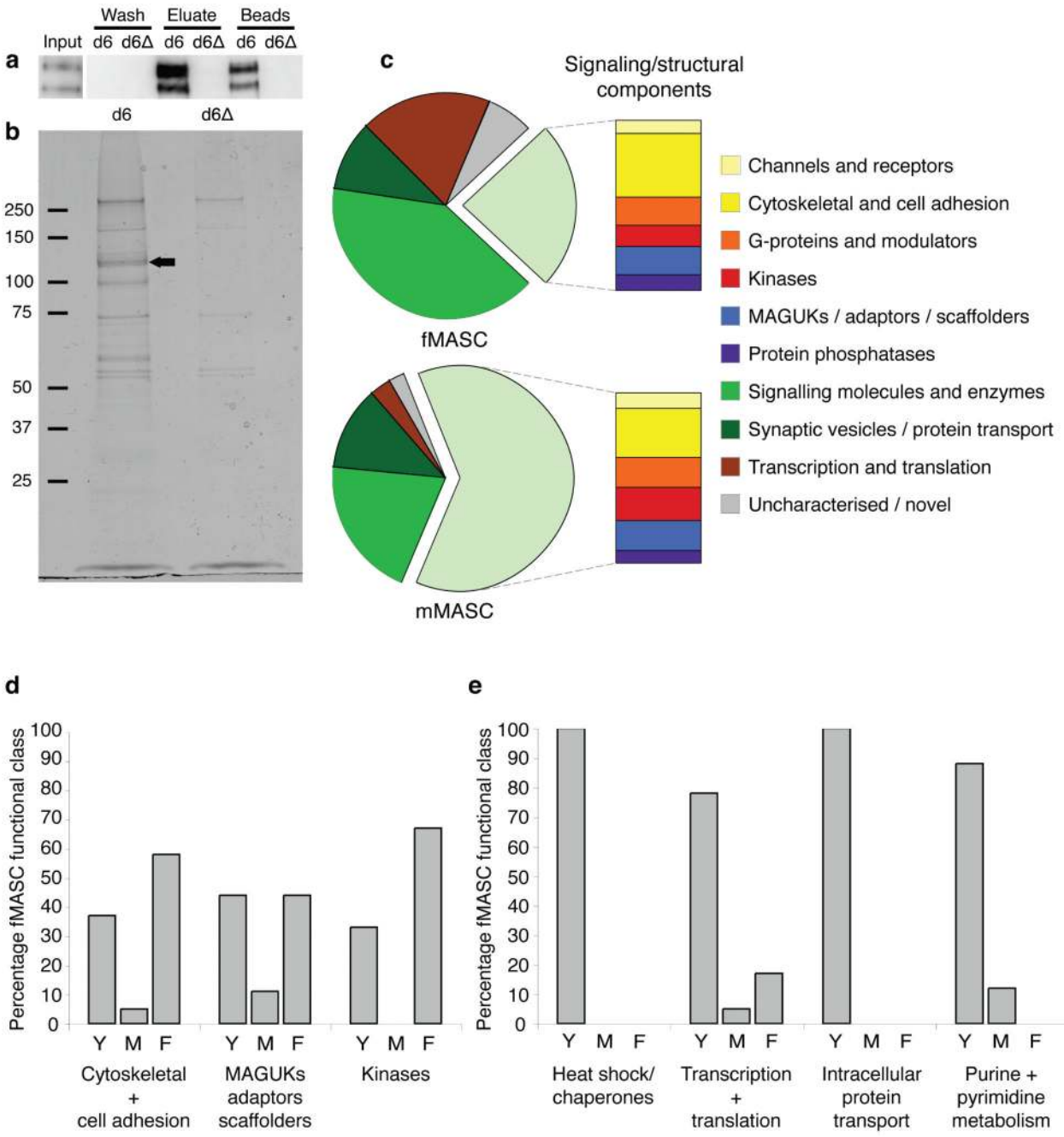

Figure 2. Proteomic analysis of the *Drosophila* MASC

a) DLG binds *Drosophila* NR2 C-terminal peptide. Protein extracts (input) of fly head were incubated with dPEP6 (d6) or dPEP6ΔVL (d6Δ) resin immobilised peptide. After binding the resin was washed (wash) and bound protein eluted (eluate) and residual bound protein (beads) were assayed for the presence of *Drosophila* DLG using immunoblotting.

b) Eluates of dPEP6 (d6) and dPEP6ΔVL (d6Δ) columns were run on a 4-12% SDS PAGE gel and stained with coomassie brilliant blue. The arrow indicates the region on the gel which gave the majority of DLG peptides after mass spectrometry.

c) Pie charts show the percentage of fMASC (220 proteins) and mMASC (186 proteins) belonging to each functional protein class. Key indicates colour code for identity of specific

Europe PMC Funders Author Manuscripts

classes (downstream, effector components are encapsulated in the blue, purple and brown segments).

d) Upstream signaling components classes showing significant expansion following divergence of fly and chordate lineages. The proportion of each functional class (as a percentage of the total number of fMASC genes belonging to it) whose earliest identifiable orthologue occurs in yeast (Y), early metazoans (M, common to fly and chordate lineages) or is fly-specific (F).

e) Downstream classes are predominantly of unicellular eukaryotic origin. The proportion of each functional class (as a percentage of the total number of fMASC genes belonging to it) whose earliest identifiable orthologue occurs in yeast (Y), early metazoans (M, common to fly and chordate lineages) or is fly-specific (F).

Europe PMC Funders Author Manuscripts

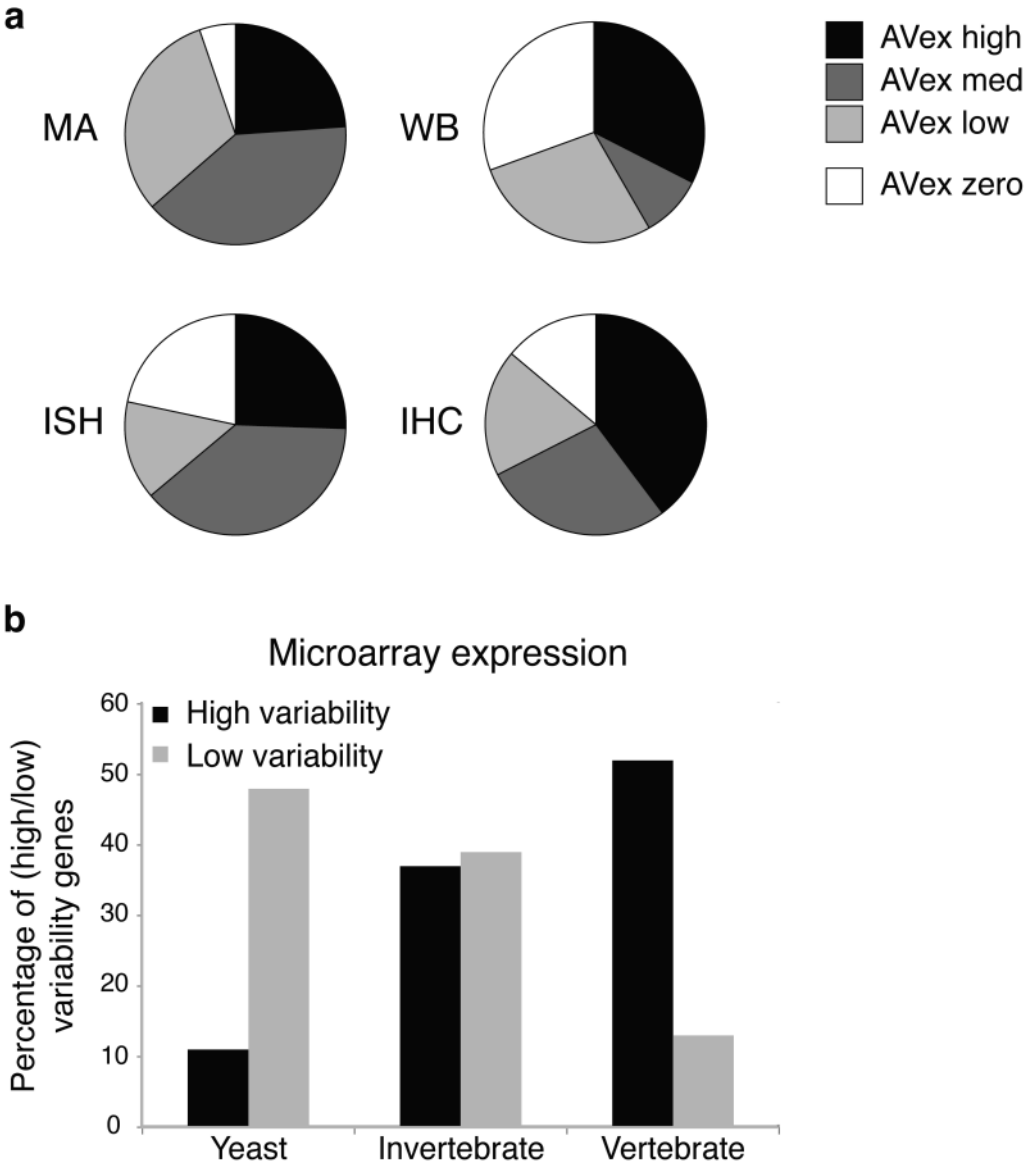

Figure 3. Variation in expression patterns in mouse brain regions

a) Relative proportions of genes with variable expression patterns using four methods. This shows the percentage of genes in each AVex class, for each of the methods of testing expression. Note that genes showing no variation in expression between brain regions ($AVex^{zero}$) are in the minority. The expression datasets are (clock-wise from top left): microarray (MA); western blot (WB); immunohistochemistry (IHC); and in-situ hybridisation (ISH). Expression variability classes are: $AVex^{zero}$ (all expression scores equal); $Avex^{low}$ (scores of only 4 and 3); $AVex^{med}$ (all scores between 4 and >1); and $AVex^{high}$ (scores of 4 and ≤1).

b) Variation in brain expression is a function of phylogeny. The graph shows the percentage of high (black) and low (grey) variability MASC genes whose earliest identifiable orthologue was present in Yeast, an Invertebrate or Vertebrate. Note that the majority of high variability genes are of vertebrate origin, and the majority of low variability genes are

of pre-metazoan origin (i.e. present in yeast). Microarray data from 22 mouse brain regions was used.

Europe PMC Funders Author Manuscripts

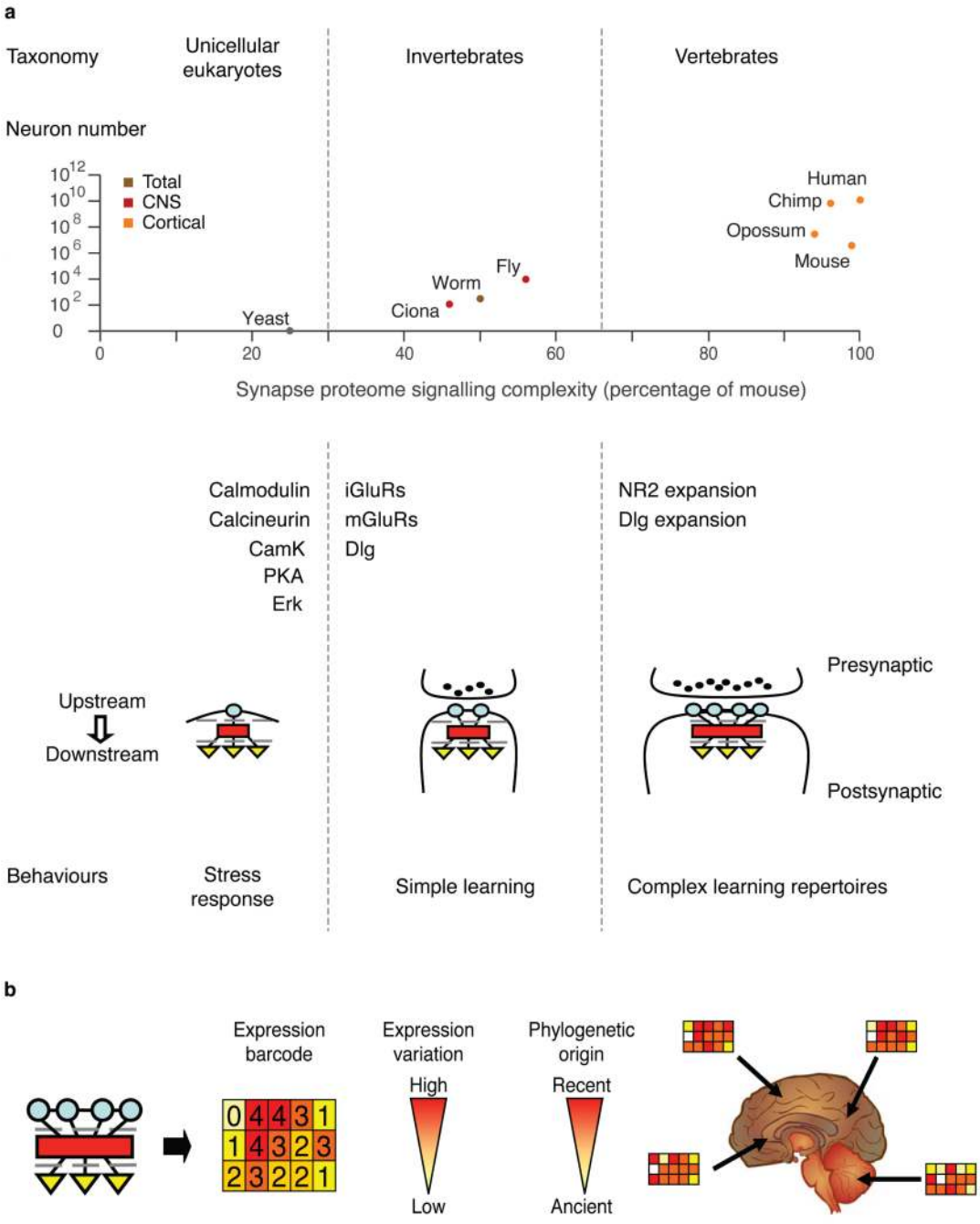

Figure 4. Summary of relationships of synapse proteome evolution with neuronal number, behaviour and expression patterns

a) Relationship of synapse/behavioural complexity to taxonomic grouping of species. In scatter plot of neuron number against synapse proteome complexity, estimates of neuron numbers were obtained from [4849503]. Synapse proteome complexity estimated as percentage of mouse MASC/PSD components possessing orthologues. Several genes of interest to learning and plasticity are listed where they first arise. The schematic representations of signaling complexes use 3 interlinked shapes (blue circles, upstream receptor/adhesion proteins; red box, signalling proteins; yellow triangle, downstream proteins). The cell membrane is indicated as a dark line and the pre- and post-synaptic terminal is indicated for

invertebrates and vertebrates. The number of blue circles and size of red box increases, illustrating their relative expansion. Behaviors indicates that while all organisms respond to their environment, ability to alter these responses and manipulate the environment show marked differences in complexity [2]. Note expansion of mammalian brain size occurs after expansion of synapse proteome complexity.

b) Mammalian MASC complexes and brain region expression variation. Schematic representation of MASC (see 4a) is shown and the expression level for 5 proteins from the 3 levels of MASC is shown (expression barcode). Upstream proteins show greater variation in expression levels and are of more recent origins. The cartoon of the brain indicates that the expression barcode is distinct for different neuronal populations.

Europe PMC Funders Author Manuscripts