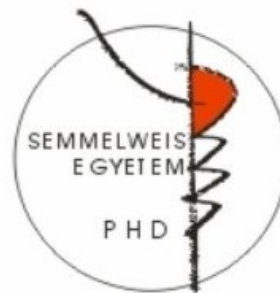


Investigation of *gap* Genes by Experimental and Network Approaches

PhD Theses

Márton Dávid Gyurkó

Semmelweis University
Doctoral School of Molecular Medicine



Supervisor: Peter Csermely, D.Sc
Csaba Sőti, MD, D.Sc

Official reviewers: János Barna, Ph.D
Miklós Cserző, Ph.D

Head of the examination committee: Barna Vásárhelyi, MD, D.Sc
Members of the examination committee: Attila Patócs, MD, Ph.D
István Miklós, Ph.D

Budapest
2015

Introduction

The neurocognitive diseases accompanied by the deterioration of learning ability and memory, and the tumor morbidities are world-wide among the most burdensome conditions for individuals and for the society as well. Though the molecular background of both clinical fields are deeply studied, their overlap is not so well known. As a part of this overlap, the GTPase activating proteins (GAPs) facilitate the intrinsic guanine-phosphatase activity of small G-proteins. The RasGAP proteins are the specific activators of the Rat sarcoma protein (Ras), which belongs to the family of small GTPases. They facilitate the GTP hydrolysing activity of Ras, and therefore inhibit the Ras/MAPK pathway.

The human SynGAP protein is the most known member of the RasGAP family in learning and memory. Its activation by Calcium/calmodulin kinase II is assumed to link the changes of intracellular calcium level, the activation of N-methyl-D-aspartate receptors, the synaptic volume changes induced by small G-proteins, the exposure of AMPA receptors and ultimately synaptic potentiation. The morbidities caused by germline mutations of the Ras/MAPK pathway, called Rasopathies, also shed light indirectly on the role of RasGAPs in human learning and memory. Their common characteristics are developmental abnormalities, increased susceptibility against tumors and cognitive symptoms, including the defects of learning. One of the most common Rasopathy involving almost 1 in 2000 fetus is neurofibromatosis type I (von Recklinghausen disease), which is caused by the mutation of neurofibromin 1, a GAP encoding gene.

Rat, mouse, fly, zebrafish and roundworms are all established model systems in learning and memory. The phylogenetic tree of *gap* genes is similarly wide, yet the *Caenorhabditis elegans* nematode is the only organism, in which not only the single, but the double mutants of the three known *gap* genes are also viable. This model system offers a wide range of behavioral and molecular biological tools to characterize learning and memory. Its short lifespan decreases the required time of experiments, and allows the examination of a large number of animals increasing statistical reliability.

Members of the *gap* gene family in *C. elegans* are the *gap-1*, *gap-2* and *gap-3* genes. They were described in the context of developmental processes as the regulators of the Ras/MPK-1 MAP kinase pathway. Their similarity to their human orthologues are 87%, 79% and 90%, respectively. The RasGAPs act through the Ras/mitogen activated protein kinase (MAPK) pathway to our best knowledge.

The untangling of intracellular RasGAP signaling in *C. elegans* poses multiple challenges. The involvement of the Ras/MAPK pathway is assumable, which, however, can be intertwined in unclear or not trivial ways with several other signaling pathways due to its central role. Network science provides useful tools to approach this complexity, enabling the intuitive understanding of the rich structure of molecular pathways potentially explaining the observed phenotypes. It is also a challenge, that according to my prior studies, the signaling pathways in *C. elegans* are much less explored compared to the most investigated human counterparts. The pathways of learning, memory and the Ras/MAPK signaling, however, are evolutionary well conserved. Therefore, network science and bioinformatics can be called on again: the missing pieces of *C. elegans* signaling can be filled in with elements of human signaling. Another challenge is the fragmentation of these data among databases and numerous different nomenclatures, e.g. that the same molecules or biological processes are referred to by different names. Large scale data integration is the solution to these challenges, the synthesized unification of the sources by the same nomenclature, which is a proven systems biological and pharmaceutical approach.

Objectives

1. The first goal of my research was the comprehensive experimental phenotypization of the loss of function *gap* mutations regarding learning and memory in *C. elegans* model system. This included the study of chemotaxis, motility, learning, short- and long-term associative memory, the involvement of Ras in these processes, and the setup of the experimental methods.
2. The second goal of my doctoral work was defined as the large scale integration and unification of the available data of intracellular RasGAP signaling. This also covered data quality improvements.

3. The third goal was to create the small-scale network of the Ras/MAPK pathway and the most important pathways involved in learning and memory. This was based on the integrated data, was neuron-specific, small scale and manually curated. The resulting network represents the molecular signaling of the RasGAPs in physiological and pathological processes, as well as their pharmaceutical relations.

Methods

Strains. The strains carrying the following mutations were used to study the role of GAP proteins in learning and memory: *gap-1(gal33)*, *gap-2(tm748)*, *gap-3(gal39)*, *gap-1(gal33) gap-2(tm748)*, *gap-1(gal33);gap-3(gal39)*, *gap-2(tm748);gap-3(gal39)* loss of function *gap* mutants, *eri-1(mg366);lin-15B(n744)* (KP3948) RNA-sensitive *C. elegans* strain. Learning and memory assays were performed using *gap-x;let-60(n2021)* double mutants to study the role of Ras in these processes. *C. elegans* N2 Bristol variant served as control otherwise noted. The strains were maintained in incubators at permanent 20° Celsius in 10 cm Petri dishes containing 20 ml NGM (Nematode Growth Medium).

Chemotaxis assay. To characterize the smelling of worms, the inherent attraction to volatile compounds, namely diacetyl, benzaldehyde and isoamyl alcohol, was quantified. 50-200 animals were dropped in the middle of the test plates, and the number of worms were counted at both sides and outside these areas. The chemotaxis index was defined as:

$$\text{Chemotaxis index (CI)} = (A-B) / (A+B+C)$$

where A is the number of worms found on the area penetrated with the attractant, B denotes the number of animals on the control area, and C is the number of worms outside these areas. The neurobehavioral methods were established in collaboration with the Division of Molecular Neuroscience, Department of Psychology, University of Basel.

Motility assays. The number of body bends per minute are recorded during motility assays in baseline feeding activity, food searching activity and feeding following 1 hour of starvation.

Learning and short term associative memory assays. In these experiments the above described inherent attraction towards diacetyl is associated with a negative stimulus, starvation (conditioning), which leads to avoidance. The naïve, conditioned, and regenerated animals were observed in tests detailed at chemotaxis assays. The results describe the naïve characteristics, learning ability and memory function of the worms, respectively.

Long-term memory assays. The effect of mutations on long-term associative memory was investigated similarly to learning and short-term memory, although conditioning and recovery were repeated 3 times, and the attraction of animals towards diacetyl was tested immediately, 16 and 24 hours after conditioning.

Transgene *gap-1(ga133);sur-5::mDsRed* strain. The gene encoding dsRed protein, driven by *sur-5* promoter providing full body expression, was merged with *gap-1* gene, and the resulting construct was microinjected into worms with *gap-1(ga133)* genotype.

RNA interference experiments. For the RNAi experiments the RNA-sensitized *eri-1(mg366);lin-15B(n744)* (KP3948) strain was fed with bacteria carrying *gap-1*, *gap-2* and *gap-3* double stranded RNA.

Bioinformatics methods. To create ComPPI, the following softwares were used: nginx web server running on Ubuntu 14.04 linux, MySQL 5 database manager, PHP 5, Javascript 1.2, HTML 4, CSS 3 and Python 3.4 programming languages, Symfony 2 and jQuery 1.4 frameworks. Network visualization was achieved using the d3.js library, Cytoscape 3 and GIMP 2.8 softwares.

Statistical analysis, graphs and figures. I wrote a custom program in Python 3.4 to statistically analyze and visualize the experimental results. This used the numeric library numpy 1.8, the graphical library matplotlib 1.4, and the statistical module implemented in SciPy 0.13.3. Significance was calculated with Welch's t-test and type II ANOVA, p values always refer to two-tailed t-tests. Multiple populations were always compared using Bonferroni-correction.

Results

The effects of *gap* loss of function on *C. elegans* chemotaxis. Learning and memory assays depend on chemotaxis, because worms acquire the environmental clues by olfaction. Chemotaxis tests therefore targeted the attraction of loss of function *gap* mutants towards diacetyl, benzaldehyde and isoamyl alcohol.

Every *gap* mutant strain reacted to olfactory clues. The *gap-2(tm748)*, *gap-3(ga139)*, *gap-1(ga133)* *gap-2(tm748)* and *gap-2(tm748);gap-3(ga139)* mutants had similar chemotaxis response to low and high concentration of diacetyl, benzaldehyde and isoamyl alcohol as wild type. The *gap-1(ga133)* mutants, however, had decreased attraction towards high concentration of diacetyl and both concentrations of isoamyl alcohol, while the attraction to

benzaldehyde remained unaffected. Decreased attraction to high concentration of diacetyl was found for the *gap-3(gal139)* strain too, which was not observable in any other experimental condition. The *gap-1(gal133);gap-3(gal139)* mutants had chemosensory defect towards all used attractants. Except this strain, the *gap* mutations did not cause significant changes in the sensation of low concentration of diacetyl, therefore they did not affect the negative associative learning and memory assays relying on olfaction.

The *gap* mutations do not affect the motility of worms. Similar to olfaction, motility defects may affect behavioral results, therefore it is necessary to test the motor function of mutant worms. This was quantified as the number of body bends per minute in baseline feeding activity, food searching activity and feeding activity after 1 hour of starvation. The body bends for baseline feeding activity and feeding after starvation were the same as for the N2 wild type. The same holds for food searching activity except for the significantly decreased motility of the *gap-1(gal133);gap-3(gal139)* strain. A possible explanation of this defect is the deteriorated motivation rooting in chemosensory defect, because no such difference was observable in baseline feeding activity and feeding after starvation.

The motor system and its neuronal control is therefore not affected in any experimental condition by the *gap-1(gal133)*, *gap-2(tm748)*, *gap-3(gal139)*, *gap-1(gal133) gap-2(tm748)*, *gap-2(tm748);gap-3(gal139)* mutations, whereas the *gap-1(gal133);gap-3(gal139)* strain was excluded from further experiments due to its chemosensory and potential locomotor defect.

The complex interplay of RasGAPs is required for learning and short-term associative memory. During learning and short-term associative memory assays the inherent attraction of worms to diacetyl is associated with the negative stimulus of starvation, therefore an aversive behavior arises, which can describe the learning ability of the animals. After half an hour recovery the remaining conditioned population is tested in a chemotaxis assay to characterize memory function.

Conditioning dramatically decreased the attraction of worms to diacetyl in every case. The *gap-1(gal133)* strain, however, showed significant learning defect compared to the wild type. Memory defect was measurable too, although the type II ANOVA test comparing the conditioned and recovery phase of the mutant animals revealed that this defect was the consequence of deteriorated learning. The *gap-2(tm748)*, *gap-3(gal139)*, *gap-1(gal133) gap-2(tm748)* and *gap-2(tm748);gap-3(gal139)* double mutant strains all had short-term memory defect without learning defect. The *gap-1(gal133);gap-3(gal139)* double mutant strain was not

assessable due to its chemosensory defect. In summary, the results show that GAP-1 is mainly involved in learning, while GAP-2 and GAP-3 play a role in the molecular processes of short-term memory formation.

RasGAPs are involved in the formation of long-term associative memory. The principle of long-term associative memory assays is similar to those of short-term, except that the conditioning phase together with the following recovery phase were repeated three times, and the animals were tested after 16 and 24 hours too. The earlier observed learning defect of the *gap-1(ga133)* strain was not measurable, which phenomenon may be explained by the repeated conditioning. The *gap-2(tm748)*, *gap-3(ga139)*, *gap-1(ga133) gap-2(tm748)* and *gap-2(tm748);gap-3(ga139)* strains all had significant long-term associative memory defect, while *gap-2(tm748);gap-3(ga139)* also showed learning defect.

Validation of the roles of *gap* genes with gene silencing and rescue lines. The relation between mutant phenotypes and *gap* genes was confirmed by gene silencing with RNA interference and by rescued *gap* lines.

The *let-60* gene is necessary for the observed loss of function *gap* phenotypes. The *gap* mutations and the hypofunctional *let-60* mutation, encoding the LET-60 *C. elegans* Ras protein, were combined and tested too, because the *gap* mutation leads to the hyperfunction of LET-60 Ras, which may be compensated by the hypofunction of *let-60*. The results indeed indicate that the *gap-1(ga133);let-60(n2021)*, *gap-2(tm748);let-60(n2021)* and *gap-3(ga139);let-60(n2021)* double mutant strains had no learning or memory defect. These altogether suggest that the balance of Ras/MAPK pathway is required for appropriate learning and memory formation.

The experimental results are summarized in Table 1.

Table 1. The phenotypes observed for *gap* mutations.

Phenotypes identical to wild type are denoted by 'OK', mutant phenotypes are marked by 'Defect' regarding the sensory system (Sensorium), locomotor system (Motility), learning, short-term associative memory (STAM) and long-term associative memory (LTAM).

	Sensorium	Motility	Learning	STAM	LTAM
<i>gap-1</i>	OK	OK	Defect	OK	OK
<i>gap-2</i>	OK	OK	OK	Defect	Defect
<i>gap-3</i>	OK	OK	OK	Defect	Defect
<i>gap-1 gap-2</i>	OK	OK	OK	Defect	Defect
<i>gap-1;gap-3</i>	Defect	OK*	N/V	N/V	N/V

<i>gap-2;gap-3</i>	OK	OK	Defect	Defect	Defect
--------------------	----	----	--------	--------	--------

* The motility defect observed for the *gap-1(ga133lf);gap-3(ga139lf)* mutant strain during food searching activity is explained by the chemosensory defect.

ComPPI, the compartmentalized protein-protein interaction database. ComPPI is an integrated database containing proteins, their interactions and subcellular localizations in a network scientific manner, and is also a software package to analyze and serve these data.

High throughput experiments produce huge amount of data, but only a fraction of these are available in public databases. The data are recorded in different nomenclatures and scattered among various sources. The overlap between these sources is low, therefore a comprehensive search is time consuming. ComPPI is our solution to these challenges. Its main components are the database storing all data, the so called interfaces that are responsible for loading the data from source databases, and the website that allows the search, download and display of these data. The software package also contains source code to analyze the data, and infrastructural components to handle and install the whole package.

ComPPI integrates and synchronizes data from 9 protein-protein interaction and 8 subcellular localization databases. The result is a dataset containing 83,753 proteins, 1,059,650 interactions and 195,815 localizations, leading to the largest publicly available protein-protein interaction database. The biological likelihood of interactions and localizations, e.g. the reliability of data is described by separate confidence scores.

The ComPPI website provides a user friendly search form to support small scale data analysis. The most important details of proteins, their interactions and intracellular localizations can be browsed, filtered and displayed even without bioinformatics skills. The result sets can be exported too. Usability is further improved by local helps and tutorials.

The whole database and its previous versions can be downloaded, which serves large scale data analysis and scientific reproducibility. From these datasets and the source code the full software package can be rebuilt with relevant proficiency, and can be run even on a common desktop computer.

The neighborhood network of RasGAPs in ComPPI. The protein-protein interaction neighborhood of RasGAP family proteins includes RASA1, RASA2, RASA3, SynGAP1, CAPRI, RASAL1, RASAL2 and RASAL3 proteins, their first neighbors and their interactions. The whole network forms a single component (193 nodes, 1365 edges), which segregates to 5 components on the removal of polyubiquitin C precursor. Signaling functions,

neurogenesis, axonogenesis, and functions related to Ras signaling can be found among the biological roles of the members of this network. Regarding subcellular localizations, most of the proteins in the neighborhood network are anchored to cell membrane, or located in cytosol or nucleus.

The network of Ras/MAPK, IP₃/DAG/PKC, cAMP/PKA, Ras/PI₃K pathways and Ca²⁺ signaling. Based on the RasGAP neighborhood network from ComPPI and on the pathway data from Kyoto Encyclopedia of Genes and Genomes, I created the human orthologome, e.g. network of the Ras/MAPK pathway and the major pathways related to learning and memory. The network was integrated using the UniProt nomenclature, and I verified manually for each node and link that publications and neighbors should allow the neuronal presence of the molecule.

This manually curated network core contained 55 nodes and 77 links, its average node degree is 2.836, the average path length is 3.611, the diameter is 7, and the network does not contain segregated node. There were cross-talks among all included pathways. Between the Ras/MAPK pathway and the pathways related to learning and memory, 18 cross-talks were identified, 7 of which modulated the activity of Ras/MAPK pathway through GAPs and GEFs. Through the revealed cross-talks the Ras/MAPK pathway is connected to the regulation of NMDA and AMPA receptors, also to cytoskeletal changes, and may have a direct role in the regulation of the expression of synaptic proteins too. These processes are fairly unexplored in human cells and there are no experimental data in *C. elegans* model system.

Further noteworthy are that the Ras/MAPK pathway is involved in learning and memory, there are antitumor drugs targeting this pathway, and the adverse effects of chemotherapy overlap with the symptoms of Rasopathies. Taking these altogether, the hypothesis arises that the cognitive dysfunction caused by chemotherapeutic treatments targeting the Ras/MAPK pathway and the cognitive symptoms of Rasopathies share a common molecular mechanism.

The analysis of the RasGAP signaling network yielded Raf as the node with the highest betweenness centrality, a measure correlating well with importance in signaling. Indeed, most of the antitumor drugs related to this pathway target Raf. On the other hand, the pharmacological targeting of Ras is difficult, therefore the molecules modulating it, the RasGAPs for example, provide promising systems pharmaceutical directions.

Conclusions

My doctoral work covered the phenotypization of the *gap* gene family in *C. elegans* model system regarding learning and memory, and these results and the potential signaling of GAP proteins were put into pathological and clinical context using bioinformatics and network science.

The main novel scientific results of my doctoral thesis can be summarized as follows:

1. I proved that the *gap* genes are required in *C. elegans* for chemotaxis, learning, and the formation of short- and long-term associative memory. Mainly *gap-1* and *gap-3* are related to chemotaxis. *gap-1* also plays a major role in learning. *gap-2* and *gap-3* contribute redundantly to short- and long-term associative memory formation. I showed as well that *let-60*, the gene encoding the *C. elegans* Ras, is required for these processes.
2. We created ComPPI, the compartmentalized protein-protein interaction database. It provides details of protein interactions and subcellular localizations in an integrated, unified manner that also quantitatively describes the confidence of these data. ComPPI supports small scale, targeted protein analysis and large scale, interactome-based data analysis, too.
3. I created the unique network representing the human, neuronal signaling of RasGAPs from the Ras/MAPK, IP₃/DAG/PKC, cAMP/PKA, Ras/PI₃K pathways and Ca²⁺ signaling. 18 cross-talks were revealed among the Ras/MAPK pathway and the major pathways of learning and memory. I proposed the hypothesis of shared molecular mechanisms behind these processes and the cognitive adverse effects of specific antitumor therapies, and the potential candidate molecules bridging the Ras/MAPK pathway and neurodegenerative diseases.

Bibliography of the candidate's publications

Publications related to the subject of dissertation

Gyurkó MD, Csermely P, Soti C, Stetak A. (2015) Distinct roles of the RasGAP family proteins in *C. elegans* associative learning and memory. *Sci. Rep.* 5:15084.

Veres DV, **Gyurkó MD**, Thaler B, Szalay KZ, Fazekas D, Korcsmáros T, Csermely P. (2015) ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res* 43:D485–D493.

Gyurkó MD, Stetak A, Soti C, Csermely P. (2015) Multitarget network strategies to influence memory and forgetting: The Ras/Mapk pathway as a novel option. *Mini-Rev Med Chem* 15:1–9.

Publications not related to the subject of dissertation

Gyurkó MD, Soti C, Stetak A, Csermely P. (2014) System level mechanisms of adaptation, learning, memory formation and evolvability: the role of chaperone and other networks. *Curr Protein Pept Sci* 15:171–188.

Gyurkó MD, Veres DV, Módos D, Lenti K, Korcsmáros T, Csermely P. (2013) Adaptation and learning of molecular networks as a description of cancer development at the systems-level: Potential use in anti-cancer therapies. *Semin Cancer Biol* 23:262-269.

Simkó GI, **Gyurkó MD**, Veres D V, Nánási T, Csermely P. (2009) Network strategies to understand the aging process and help age-related drug design. *Genome Med* 1:90.