Investigation of different isoforms and genetic variants of the *SNAP-25* gene

Doctoral Theses

Nóra Németh MD

Semmelweis University
Molecular Medicine PhD School

Supervisor: Zsolt Rónai MD, Ph.D

Official reviewers: Beáta Törőcsik MD, Ph.D
Krisztián Buza Ph.D

Head of the Final Examination Committe: László Tretter MD, Ds.C
Members of the Final Examination Committe: István Voszka Ph.D
Krisztina Vellainé Takács Ph.D

Budapest
2016

Introduction

The development of complex diseases (like diabetes, different types of cancer or several psychiatric disorders), is known to be caused by genetic and environmental factors, albeit the ratio of these components can significantly differ in case of different phenotypes and illnesses. The appearance of the symptoms is the result of the interaction of these ethological factors, and the role of the individual components is small, and thus often challenging to demonstrate and identify. Consequently identification of the genetic background of complex diseases is rather difficult, despite this knowledge is fundamental for the development of causal therapy and the effective primary and secondary prevention.

The research presented here investigated the different isoforms and genetic variants of the *SNAP-25* gene. The protein coded by this gene participates in generation of the SNARE complex, thus – together with other proteins – it plays an important role in the exocytosis of intracellular vesicles. The *SNAP-25* gene encodes two different transcriptional variants and these isoforms generates two distinct proteins. The two mRNA variants differ in exon 5, which results in 9 different aminoacids at the protein level. It was observed, that the total amount as well as the ratio of the two isoforms changed during the ontogeny, moreover the alteration of the expression of the isoforms was demonstrated in different diseases, consequently the quick and accurate determination of the two variants is of significant practical as well as theoretical importance.

Several SNPs of the *SNAP-25* gene were demonstrated to be the genetic components of different psychiatric disorders. Polymorphisms in the non-coding, regulatory regions (promoter, 5’ and 3’ UTR) might be of special interest, because these loci can influ-

ence the amount of the generated protein. It is long well known, that the promoter SNPs might result in the modulation of transcription by altering the binding efficiency of transcription factors. More and more attention has recently focused on the regulatory mechanisms implemented by the micro-RNAs. More than 2500 matured micro-RNAs have been identified so far, and their importance is shown by the fact, that more than half of the transcriptom is regulated by micro-RNAs. The complexity of the system is also represented by the observation, that most of the 3' UTRs of mRNAs bind numerous micro-RNAs, and also vica versa: most microRNAs take part in the regulation of the translation of several proteins. The picture gets even more complex when keeping in mind, that although this regulation is carried out by RNA–RNA interaction, despite simple sequence analysis (identification of complementary nucleotides) can not unambiguously identify, which mRNA is targeted by a given micro-RNA. This complex regulatory network is further modulated by the miR-SNPs, which are polymorphisms in the 3' UTR of mRNAs altering the binding efficiency of a given microRNA. One of the goals of our research was the investigation of such miRSNPs.

Analysis of a miR-SNP can be carried out from many different points of view. Genetic background of complex diseases is often investigated by association analysis, which studies the putative joint occurrence of a given genetic constellation (allele, genotype, haplotype) and the chosen phenotype of interest. Statistical verification, however, does not prove the causal connection between the genetic variant and the phenotype. *In vitro* functional analyses (i.e. luciferase reporter system) can help to discover the role of a given SNP in the biological regulation, which can underpin the result of the association analysis. The current work focused on the analysis of selected SNPs of the *SNAP-25* gene by this principle.

Objectives

The main focus of the current research was the **investigation of different isoforms and selected genetic polymorphisms of the *SNAP-25* gene**. Specific aims of the current work were the following:

1. **Specific determination of the expression level of the two different *SNAP-25* gene isoforms in 10 human tissues**. For the quantitative analysis of the isoforms we aimed **to elaborate two, independent methods** in order to improve reliability. For optimization and validation of the two techniques, recombinant DNA-**constructs were generated, that express the two different isoforms**, respectively.

2. **Psychogenetic association study** was planned to investigate the putative association between the **3' and 5' UTR SNPs of the *SNAP-25* gene** and the **impulsivity personal trait** in healthy adult persons.

3. We aimed to use molecular biologic approaches to analyze, whether the two **SNPs** (rs3746544, rs1051312) **in the 3' UTR** of the *SNAP-25* gene, localized in close proximity (4 bp distance) to each other, **influence the binding efficiency of miR-610** and consequently the **amount of the generated protein**.

Methods

1. Investigation of the transcript isoforms of the *SNAP-25* gene

Total RNA of different human tissues was used for downstream cDNA synthesis, followed by the amplification of the coding region of the *SNAP-25* gene. The generated insert was sub-cloned into pcDNA3.1(–) vector using standard protocols. Base sequence of the generated DNA-constructs was confirmed by direct sequencing.

For the specific detection of the two isoforms of the *SNAP-25* gene two different methods were elaborated. The first technique applied PCR-RFLP, whereas the second one was based on real-time quantitative PCR. These Methods were validated by the DNA-constructs mentioned above.

Quantitative analysis using the PCR-RFLP method was carried out by subsequent horizontal agarose gel electrophoresis followed by densitometry. For higher efficiency and reliability capillary gel electrophoresis was also applied for the detection of the digested PCR-products.

2. Psychogenetic association study analyzing the genetic variants of the *SNAP-25* gene

901 healthy young adults participated in this study, which was conducted in collaboration with the Psychogenetic workgroup of the ELTE University. The participants were informed about the goals and the methods of the study and they signed an informed consent. Participants provided buccal samples for genetic analyses. Research protocol was approved by the Local Ethical Committee (ETT-TUKEB, Scientific and Research Ethics Committee of the Medical Research Council). The impulsivity personality trait was measured by the 11th version of the Barratt Impulsivity Scale. After reviewing the relevant literature and analyzing the data from our *in silico* research for the genetic analysis we chose 4 SNPs from the regulating regions of the *SNAP-25* gene – 5’

(rs6077690 A/T, rs6039769 A/C) and 3’ (rs1051312 C/T, rs3746544 G/T). From the buccal samples DNA was isolated, genotyping was executed by PCR-RFLP or real time PCR methods. Haplotyping of the 3’ UTR SNPs could directly be done with the real time PCR technique, because of close proximity of the two investigated loci. **Linkage analysis** between these 2 SNPs and the 5’ UTR polymorphisms was carried out by the HaploView 4.2 software. Association of the genotype and haplotype results and of the impulsivity personality trait was assessed by ANOVA statistical method.

3. *In vitro* analysis of the putative functional role of the *SNAP-25* gene variants

Regarding the results of our association study we intended to further analyze the two genetic variants (rs1051312, rs3746544) of the 3’ UTR region. The whole 3’ UTR region of the SNAP-25 gene was subcloned into pMIR-REPORT vector, the theoretically possible haplotypes of the 2 SNPs were constructed with site-directed mutagenesis. These reporter constructs together with the miR-641 micro-RNA were co-transfected in HEK293T cells, a construct coding for β-galactosidase was applied as internal control.

Cells were harvested 24 hours after transgfection extracted by freeze–thaw cycles and subsequent centrifugation. Enzyme activities (luciferase and beta galactosidase) were subsequently measured by luminometry and photometry.

Abbreviations

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase, miR: microRNA, miRSNP: SNPs influencing the binding of micro-RNA, PCR: polymerase chain reaction, RFLP: restriction fragment length polymorphism, SNAP-25: synaptosomal-associated protein-25, SNP: single nucleotid polymorphism, UTR: untranslated region

Results

1. Investigation of the isoforms of the *SNAP-25* gene

Previous studies demonstrated, that the two isoforms of the SNAP-25 gene differ only in their 5th exon, resulting in the alteration of 9 amino acids at protein level. **Two independent methods** were elaborated as a first step of our work for the **quantitative** measurement of **the two mRNA isoforms** of the *SNAP-25* gene. For the optimization of the two different methods internal controls were prepared from pcDNA3.1(–) expression vectors, which contained the coding sequence of the SNAP-25 *a* and *b* isoforms.

The first method relied on an initial **PCR** followed by **selective digestion** of the *a* isoform and a subsequent **electrophoretic separation** and **quantitative** detection of the transcriptional variants. In the first step the appropriate region of the 5th exon was amplified by PCR. Special care was required for the optimization of cycle number during the reaction, because the amount of SNAP-25 isoforms was higher in the tissues from the central nervous system by three to four orders of magnitudes. The *Dde* I restriction endonuclease selectively digested the amplicons generated from the *a* isoform, providing the possibility of unambiguous separation and specific identification of the two transcription variants.

Our second method employed **real time PCR** technique. Locus specific primers were used for the selective amplification of the two isoforms, analysis of GAPDH expression was used as internal control. Relative amount of the two isoforms was calculated by ΔC_T method. Serially diluted samples were used to verify our techniques and to demonstrate, that the two methods can reliably be used to determine the amount of the two transcriptional variants.

After optimization of our techniques, expression levels of the two isoforms were measured in 10 different human tissues (4 of them originated from the central nervous system, whereas 6 tissues

were extra-neural). Adequate agreement of the data obtained by the real time PCR approach and by the semi-quantitative densitometry following the PCR-RFLP method could be observed.
The amount of **the *b* isoform was higher in all neural tissues**, the **highest** amounts were measured in the **frontal cortex**. Interestingly **in all non-neural tissues** the ***a* isoform was more dominant**, but the absolute amounts of both isoforms were on avarage by two orders of magnitudes lower.

2. Analysis of the SNPs of regulating regions of the *SNAP-25* gene

We investigated selected SNPs in the regulatory regions of the *SNAP-25* gene in the second part of our work. **Genetic association study** was carried out to analyze, whether there is a connection between the S**NAP-25 polymorphisms** and the **impulsivity** personality trait.
First, 5’ and 3’ regulatory regions of the *SNAP-25* gene were analyzed *in silico* to identify loci with putative functional role. Two polymorphisms from the 5’ (rs6077690, rs6039769) and from the 3’ region (rs1051312, rs3746544) were selected for downstream analysis, respectively.
901 healthy adults participated in the study. Assessing the **Hardy–Weinberg equilibrium** for genotype distributions, **no significant difference could be observed between the measured and expected genotype frequencies** in our population for any of the 4 investigated SNPs. Minor allel frequencies (MAF) were shown to be between 0.255 and 0.417.
The investigated SNPs in the 5’ UTR region were localized far away from each other on the chromosome, thus haplotype frequencies were calculated by the “regular” method. The two investigated SNPs of the **3’ UTR** region were however in close proximity (only 4 basepairs from each other), thus **direct, molecular haplotype determination** of these SNPs was possible.

Results confirmed the previous findings of our group, that the **G–C haplotype** of the rs1051312 and rs3746544 loci is **not present** in the investigated Caucasian population.
Interestingly similar results were found, when analyzing the haplotype frequencies of the 5' UTR SNPs (rs6077690, rs6039769). Although all four haplotypes existed in our population, the frequency of the T–A haplotype (0.006) was remarkably lower, than expected based on calculation using the allele frequencies of the two loci (0.161). This special condition was demonstrated by a remarkably high Lewontin's *D'* value associating with a relatively low R^2 measure.
No linkage could be observed between the investigated polymorphisms in the 3' and the 5' region.
Genotype and haplotype data were then used for the association analysis, where the putative connection between the ***SNAP-25* SNPs** and **the impulsivity endophenotype** was analyzed. Specifically, average impulsivity scores of participants with certain genotypes or haplotypes were compared to assess genotype–phenotype correlation.
We **could not find any connection** between the impulsivity and genotypes of the investigated SNPs in the **5' region** of the gene.
The rs1051312 polymorphism in the 3' UTR showed significant nominal association with the total score of the Barratt impulsivity test ($p = 0.042$), however this result did not remain significant applying the Bonferroni correction for multiple testing. Data analysis was continued by testing the haplotypes of the investigated loci. None of the haplotypes in the 5' UTR showed an association with the impulsivity endophenotype. It was detected on the other hand, that participants possessing the **T–T haplotype** of the rs1051312 and rs3746544 SNPs in the **3' UTR** achieved **significantly lower total scores** on the Barratt impulsivity scale. This effect was significant when analyzing each haplotype group individually, ($p = 0.009$) as well as when comparing participants possessing and lacking the T–T haplotype ($p = 0.003$).

3. Functional analysis of the 3’ UTR genetic variants of the *SNAP-25* gene

SNPs of the 3’ UTR region were further analyzed using molecular biological methods based on the results of the association study. The *in silico* sequence analysis revealed, that both investigated genetic variants in the 3’ UTR (rs1051312, rs3746544) might influence the interaction between miR-641 and the mRNA of SNAP-25, as they are both localized in the *seed* region of the target sequence, recognized by the micro-RNA on the mRNA.

First we investigated the expression level of SNAP-25 and miR-641 in the HEK-293T cell-line. Are results were in agreement with previous findings, that SNAP-25 is present in this cell type. MiR-641 was also successfully detected, however its level was rather low compared to miR-196b used as internal control.

Next, the whole **3’ UTR of the *SNAP-25*** gene was cloned into pMIR-REPORT **luciferase vector** and the four, theoretically possible haplotype versions were constructed by site directed mutagenesis. A series of dilutions were tested to find the optimal amounts of constructs used in the transient transfection experiments.

We analyzed the assumed effects of rs1051312 and rs3746544 SNPs: **the relative luciferase activity values** of the constructs containing the four haplotypes were compared. Lowest activity was detected for the **T–T** haplotype, in this case the sequences of the miR-641 and the 3’ UTR region of the SNAP-25 (in the reporter vector construct) are perfectly complementary to each other. The relative luciferase activities of the constructs containing the G–T and T–C haplotypes were 1.8 and 2.1 times higher, respectively compared to the T–T variant. These results showed that one nucleotide change significantly elevated the activity of the luciferase level, however the difference between G–T and T–C was not significant, suggesting that the localization of the alteration inside the seed sequence does not play a crucial role in the miRNA binding.

In the case of the G–C haplotype 2 nucleotides are not complementary with each other, which resulted in a 4.6 times elevateon of the relative luciferase level.

Conclusions

The rising prevalence of complex diseases (like for example type 2 diabetes, high blood pressure or diverse psychiatric diseases) and the still incomplete knowledge about the exact pathophysiological mechanisms of the development of these illnesses give the importance and actuality of investigations of the genetic and enviromental background of complex diseases.

Observations prove that complex diseases also show significant familiar accumulation, though the explanation of these illnesses by only one or a few genes like in the pathomechanisms of monogenetic diseases is not possible. Numerous genetic factors and most different enviromental factors as well as the complex interactions of all these elements are responsible for the developement of complex diseases.

SNAP-25, the gene investigated and presented in this work is an excellent example for demonstrating that genetic variations of the coding region are solely not explaining the variability of proteins in physiological or in pathophysiological conditions.

The wide variety of protein isoforms proposes, that proteins have far more functions than known today. For example the role of SNAP-25 outside the central nervous system is highly probable, for example in secretory organs, like the pancreas.

Different isoforms and genetic variants in the non-coding regions play a major role in the quality and quantity of the proteome of a cell. Changes in the amount of different isoforms, a switch in the isoform-proportions can be sensitive biomarkers of various pathophysiological conditions.

Optimal amount of proteins, precisely controlled in time and space are of crucial importance in the life of a cell or organism. MicroRNAs play an essential role in the fine tuning of this system, which also makes them good candidates for diagnostics and even for therapy.

Our results support the fact, that not only the microRNAs themselves but also the genetic variations in their binding regions on

the target mRNAs have a relevant impact on the optimal functioning or disfunctioning of the translation process.
In contrast to monogenetic diseases genotype–phenotype correlations of complex disorders are far more compound and thus not always straightforward. Accordingly by defining more and more components of the heterogeneous genetic background of complex diseases does not only provide important new knowledge about the pathophysiology of these diseases, but might also contribute to the elaboration of more effective prevention strategies, new diagnostic tools or even developing new, causal therapies for complex diseases.

Bibliography of the candidate's publications

Publications related to the theses

1. Németh N, Kerékgyártó M, Guttmann A, Rónai Z. (2013) Rapid identification of human SNAP-25 transcript variants by a miniatureized capillary electrophoresis system. Electrophoresis, 35:379-84

2. Németh N, Kovács-Nagy R, Székely A, Sasvári-Székely M, Rónai Z. (2013) Association of impulsivity and polymorphic microRNA-641 target sites in the SNAP-25 gene. PLoS One, 8:e84207

Publications unrelated of the theses

1. Nagy G, Kovacs-Nagy R, Kereszturi E, Somogyi A, Szekely A, Nemeth N, Hosszufalusi N, Panczel P, Ronai Z, Sasvari-Szekely M. (2009) Association of hypoxia inducible factor-1 alpha gene polymorphism with both type 1 and type 2 diabetes in a Caucasian (Hungarian) sample. BMC Med. Gen., 10:79

2. Spiró Z, Arslan MA, Somogyvári M, Nguyen MT, Smolders A, Dancsó B, Németh N, Elek Z, Braeckman BP, Csermely P, Sőti C. (2012) RNA interference links oxidative stress to the inhibition of heat stress adaptation. Antioxid. Redox Signal., 17:890-901

3. Kerékgyártó M, Németh N, Kerekes T, Rónai Z, Guttman A. (2013) Ultrafast haplotyping of putative microRNA-binding sites in the WFS1 gene by multiplex polymerase chain reaction and capillary gel electrophoresis. J Chromatogr A, 1286:229-34

4. Eszter Kotyuk, Gergely Keszler, Nora Nemeth, Zsolt Ronai, Maria Sasvari-Szekely, Anna Szekely. (2013) Glial cell line-

derived neurotrophic factor (GDNF) as a novel candidate gene of anxiety. PLOS One, 8:e80613

5. Kótyuk Eszter, Németh Nóra, Halmai Zsuzsa, Faludi Gábor, Sasvári-Székely Mária, Székely Anna. (2013) A hangulati dimenziók és a glia-eredetű növekedési faktort kódoló gén polimorfizmusainak összefüggése depresszióval diagnosztizált mintán. Neuropsychopharmacol Hung, 15:63-72
6. Elek Z, Németh N, Nagy G, Németh H, Somogyi A, Hosszufalusi N, Sasvári-Székely M, Rónai Z. (2016) MicroRNA Binding Site Polymorphisms in the WFS1 Gene Are Risk factors of diabetes mellitus. PLoS One, 10:e0139519

Acknowledgement

First of all I would like to thank my supervisior, Zsolt Rónai MD PhD. It's not measurable how much help, advise and last, but not least nice and cheering words I've got from him in the past almost 10 years. I thank him, that he taught me with unique patience all time, showed me everything, from the very beginnings: how to use a pipette correctly, through the planning and conducting of experiments, as well as how presenting our results. Later as his PhD student it was always a pleasure to work with him, he gave me always good suggestions, supported me.

I would like to thank József Mandl DSc, the head of the Pathobiochemistry PhD school and Gábor Bánhegyi, DSc, the director of our institute, the Departement of Medical Chemistry, Molecular Biology and Pathobiochemistry.

I'm also very thankful to Mária Sasvári DSc, who was once my practice leader and awakened my interest for the molecular biology. She always supported my work and helped also with her suggestions in her laboratory.

I thank my colleagues, especially Zsuzsanna Elek and Melinda Bence for their technical support, the good atmosphere during our common work and the many merry moments together.

I would like to thank our collaborators, András Guttman DSc, Márta Kerékgyártó PhD from the Debrecen University, Anna Székely PhD and Eszter Kótyuk PhD from the ELTE University and of course all participants of the study.

Finally I thank my family and friends, for all of their support and love.