Analysis of mice brain and SH-SY5Y neuroblastoma cell proteome by quantitative two-dimensional electrophoresis

Ph.D. thesis

István Földi

Department of Medical Chemistry

University of Szeged

Supervisors:

Prof. Botond Penke

Dr. Tamás Janáky

Szeged

2011.

List of publications

Original full papers, directly related to the subject of the thesis:

Charaterisation of the variation of mouse brain proteome by two-dimensional electrophoresis

Földi I, Müller G, Penke B, Janáky T

J Proteomics (2011), **74**:894-901 doi:10.1016/p.prot.2011.03.006 (IF: 3.851)

Proteomic study of the toxic effect of oligomeric Aβ1-42 *in situ* prepared from 'iso-Aβ1-42'

Földi I, Datki ZL, Szabó Z, Bozsó Z, Penke B, Janáky T

J Neurochem (2011), **117**:691-702, doi:10.1111/j.1471-4159.2011.07238.x (IF: 3.999)

Review article, directly related to the thesis:

Systems biology of Alzheimer's disease: How diverse molecular changes result in memory impairment in AD

Juhász G, Földi I, Penke B

Neurochem Int (2011), **58**:739-750, doi:10.1016/j.neuint.2011.02.008 (IF: 3.541)

Original full papers, not directly related to the subject of the thesis:

NO-sensitive guanylyl cyclase beta 1 subunit is peripherally associated to chromosomes during mitosis. Novel role in chromatin condensation and cell cycle progression

Pifarré P, Baltrons MA, Földi I, Garcia A

Int J Biochem Cell Biol (2010), **41**:1719-1730, doi:10.1016/j.biocel.2009.02.022 (IF: 4.887)

Characterisation of the interaction between amyloid-beta 1-42 and glyceraldehide phosphodehydrogenase

Verdier Y, Földi I, Sergeant N, Fülöp L, Penke Zs, Janáky T, Szűcs M, Penke B

J Pept Sci (2009), **14**:755-762, doi:10.1002/psc.998 (IF: 1.807)

Differences between normal and alpha-synuclein overexpressing SH-SY5Y neuroblastoma cells after Aβ(1-42) and NAC treatment

Hunya Á, Földi I, Szegedi V, Soós K, Zarándi M, Szabó A, Zádori D, Penke B, Datki ZL

Brain Res Bull (2009), **75**:648-654, doi:10.1016/j.brainresbull.2007.10.035 (IF: 2.184)

Introduction

Biological functions are mainly carried out by dynamic populations of proteins determined by gene expression regulation and environmental factors. The term ‘proteome’ was described as: the PROTEin complement expressed by a genOME. Proteomics is initially concerned with determining the structure, expression, localization, biochemical activity, interactions and cellular roles of as many proteins as possible. In other words, proteomics is a large-scale study of the proteome.

Quantitative proteomics aims at establishing differences in the proteome profiles from qualitative and quantitative point of view between individuals or populations, and to identify protein species which exhibit changes in expression or post-translational modifications in response to a given stimuli. In proteomics, a significant portion of the accumulated data has been obtained from quantitative two-dimensional electrophoresis (2-DE). 2-DE is the most frequently used and highest resolution technique to separate intact proteins in a single analytical run. Despite the many technical improvements this method is not automated and requires proficiency to get reliable and valid information. Thus, in my Ph.D. dissertation I put a special emphasis on quantitative 2-DE, discussing the basic principles, advantages and limitations of this technique.

Advances in proteomic technologies have equipped the field of neuroproteomics to study the proteomes of the nervous system. Neuroproteomic studies become more important and relevant, as the prevalence of some neurodegenerative diseases has increased in the last few decades. Alzheimer’s disease (AD) is the most common cause of dementia. Proteomics plays an undoubted role in understanding the molecular background of the patomechanism of AD. Proteomic analysis of clinical samples is difficult due to the fact that it is rarely possible to obtain a biopsy, thus in general brain alterations can be examined only at autopsy. Autopsy samples usually represent the end-stage of AD. Due to these difficulties, animal models and various cell lines are frequently used model systems to study the molecular background of AD progression.

Aims

Understanding and controlling the variance components in a 2-DE study are crucial. In neuroproteomics and AD research, one of the most frequently used model systems is mouse. Thus, a 2-DE analysis of mice brain proteome would be useful to estimate the magnitude of the variance components (technical and biological). Moreover, this study should reveal the quantitative limitations of this kind of experiments. Hence, I investigated the variation components of a 2-DE study analysing mice brain proteome.

In AD research, cell cultures are also frequently used model systems to investigate the molecular mechanism of the cellular events induced by different aggregation states of Aβ1-42 peptide. The exact mechanism of cell death induced by oligomeric Aβ1-42 is not elucidated, yet. Thus, a reliable 2-DE study would be useful to detect molecular changes after oligomeric Aβ1-42 treatment using SH-SY5Y neuroblastoma cells. The functional classification of the identified proteins would be helpful to understand the mechanism of oligomeric Aβ1-42 induced neuronal stress.

In the course of my Ph.D. work I attained the following aims:

1. Characterisation of the variation of mouse brain proteome by 2-DE:
   1.1. The magnitudes of the biological- and technical variation were estimated.
   1.2. Reproducibility, inter-assay studies were performed to characterise the quantitative limitations of a 2-DE study analysing mice brain proteome.
   1.3. The effect of the genetic background on the total variation was investigated analysing mouse brain proteome by 2-DE.
2. Proteomic study of the effect of oligomeric Aβ1-42 *in situ* prepared from "iso- Aβ1-42":
   2.1. Significant protein expression changes were detected after oligomeric Aβ1-42 treatment in SH-SY5Y neuroblastoma cells.
   2.2. The identified proteins were sorted into functional groups to understand which subsets of cellular proteome play a role in oligomeric Aβ1-42 induced cell stress.

Materials and methods

Experimental details for 2-DE analysis of mouse brain proteome

2-DE brain proteome profiles of littermate and non-littermate NMRI mice were analysed. To calculate the technical variance, two randomly selected samples were used for running technical replicate gel sets (n=4). To determine the total variance, seven series of biological replicate gels (four series of littermate and three series of non-littermate) were run using brain tissue taken from each mouse (n=4). Five parwise comparisons were performed to detect protein expression differences between mouse groups. Inter-experiments were made using littermate and non-littermate mice to estimate the run-to-run variance, when the animals were born and about two-month later sacrificed on the same day, and the sample preparation was performed in a short period of time. To analyse the protein expression differences between generations I compared the brain proteome profiles of two consecutive generations of the same parentage. A similar study was also made by comparing two non-littermate mice groups. Moreover, the brain proteome profiles of NMRI (outbred mouse stock) and C3H/HEN (inbred mouse strain) mice were also compared. Altogether, in this study forty gels were made and analysed.

Cell culture and oligomeric Aβ1-42 treatment

SH-SY5Y human neuroblastoma cells were grown for 10 days on 60 mm Petri-dishes. Cell differentiation was initiated by the addition of retinoic acid and phorbol ester. Before the treatment, the supernatant solution was removed and new medium was added (without fetal bovine serum and free from retinoic acid and phorbol ester). The cells were treated with oligomeric Aβ1-42 *in situ* prepared from "iso-Aβ1-42" peptide (15μM).

Sample preparation and 2-DE

2-DE analysis of mouse brain proteome

Right hemispheres of NMRI or C3H/HEN mice were homogenised in 2-DE lysis buffer supplemented with protease inhibitor. After centrifugation of the fresh homogenates

the samples were purified with 2-D Cleanup Kit. Then, the protein concentration was determined by a Non-Interfering Protein Assay Kit. A volume of samples containing 500 µg of total protein was supplemented with 2-DE lysis buffer (without protease inhibitor) and Bio-Lyte 3-10 buffer and left on 24cm, pH 3-10, NL, IPG strips for overnight rehydration. Isoelectric focusing (IEF) was performed on an IEF cell using a 24 h program for a total of 67.000 Vh. After IEF the strips were equilibrated for 2x10 min in an equilibration solution supplemented with DTE. It was followed by a second equilibration for 2x10 min with the same solution without DTE, but with IAA. After equilibration, the strips were applied to SDS-PAGE (10-14.5% gradient polyacrylamide gel, 20 x 24 cm). After SDS-PAGE, the gels were stained with RuBPs then they were scanned on a FLA-5100 laser scanner. The digitized gel images were analysed by Progenesis Samespots software.

Oligomeric Aβ1-42 treated SH-SY5Y cells

Treated and untreated cells were washed and then collected in PBS. The suspensions were centrifuged and the cell pellets were homogenised in 2-DE lysis buffer. The fresh homogenates were incubated on ice then they were lysed by sonication. The lysates were centrifuged and the supernatants were pipette into clean Eppendorf-tubes. Three samples per group were pooled to obtain a sufficient amount of protein for 2-DE (9 Petri-dishes/group, 3 gels/group). The protein concentrations were determined by a Non-Interfering Protein Assay Kit. A volume of samples containing 180 µg of total protein was supplemented with 2-DE lysis buffer and Bio-Lyte 3-10 buffer and left on 24cm, pH 3-10, NL, IPG strips for overnight rehydration. IEF was performed for a total of 75.000 Vh. Equilibration, SDS-PAGE, gel staining and scanning were performed as described above.

Statistical analysis

2-DE analysis of mouse brain proteome

The variance components were characterised by the coefficient of variation (CV%). The variation of each group were characterised by the median CV% and the CV% which encompassed 95 % of the spots (95^{th} percentile position). By making pairwise comparisons, different spots were identified. Each comparison was filtered to identify different spots having

a FDR-corrected *p*-value ≤ 0.05; power ≥ 0.8 and fold change ≥ 2.0. However the number of spots was also calculated which presented at least 1.5-fold change. Multiple testing statistical calculations (FDR and SGoF) were performed by SGoF software. Using the non-transformed, normalised spot volumes, correlation coefficient (R^2) was also calculated to measure the group-to-group variation in each pairwise comparison. Piface software was used to calculate power for the variances of littermate and non-littermate biological replicates.

Oligomeric Aβ1-42 treated SH-SY5Y cells

In 2-DE, the software-based analysis was filtered to detect different spots having *p*-value ≤ 0.05; power ≥ 0.8 and fold change ≥ 1.5. Multiple testing statistical calculations were performed by SGoF software. In Western blot analysis the differences between treated and untreated samples were determined by applying Student's T-test with a two-tailed distribution. Here a *p*-value ≤ 0.05 was considered statistically significant.

Mass-spectrometric analysis and protein identification

RuBPs stained gels were overstained with colloidal Coomassie blue G-250. Individual spots of interest were excised from the gel, destained and then subjected to in-gel digestion with trypsin. The digested protein samples were analysed on a Waters NanoAcquity UPLC system coupled with a Micromass Q-TOF premier mass spectrometer. All the obtained data were processed and peaklists were generated by Waters Proteinlynx Global server. Each MS/MS sample was analysed using a Mascot search engine. The Mascot device was set up to search in Swissprot database assuming the digestion enzyme trypsin. p*I* and Mw values of identified proteins were calculated using the Compute p*I*/Mw tool at the Expasy website (http://www.expasy.org/tools/pi_tool.html). Values of the highest Mw processed (mature) form for each protein were selected. Scaffold software was used to validate MS/MS based identifications of peptides and proteins. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least 2 identified peptides. For multiple identifications, only the protein identified with the largest amount was used for functional categorization. Protein quantity was estimated by the total number of MS/MS spectra assigned to the identified protein and average peptide ion intensity.

Quantification of Hsp70 and EEF-2 by Western blot

A volume of treated and untreated SH-SY5Y cell lysate samples containing 20 μg of total protein was separated by 12% polyacrylamide gel and transferred onto nitrocellulose membrane. The membranes were incubated for overnight with first antibodies: anti-Hsp70 (1:1000), anti-EEF2 (1:1000) anti-beta-actin (1:1000) diluted in TBS containing Tween-20 and BSA. After overnight incubation, the membranes were re-incubated with second antibodies: anti-rabbit or anti-mouse IgG (1:2500). The immune complexes were detected with chemoluminescent substrate of peroxidase by exposure to x-ray film. The films were digitised on FLA-5100 laser scanner, then the images were analysed via Multigauge software.

Results and discussion

2-DE analysis of mouse brain proteome

In present study, the effect of the genetic background on mouse brain proteome profile was investigated. Using standardised growth/housing conditions and 2-DE technical parameters, the brain proteome of littermate and non-littermate NMRI mice were analysed. Based on the first hypothesis, the genetic variability could be reduced by using littermate mice, which are born from the same parentage, so they may carry more homogenous genome. The total variation of 2-DE brain proteome profile of littermate or non-littermate NMRI and non-littermate C3H/HEN mice were estimated. Moreover, the technical variation of our 2-DE workflow was also assessed by running technical replicate gel sets. Using these data results, the proportion of technical variation in total variation was also calculated.

To determine the technical and total variation, two series of technical, four series of littermate biological and three series of non-littermate biological replicates were analysed together. In this way, the spots compared between gels had identical boundaries in all gels (36 gels). The technical variance displayed an average 9.3% median CV% and the 95% of the spots had an average CV% below 20.5%. The median CV% values of the total variation were on average 12.1% in littermate groups and 11.2% in non-littermate NMRI mouse groups. But CV% values were 27.7% ± 4.1 at the 95^{th} percentile position in littermate mouse groups and 26.8 ± 0.7 in non-littermate mouse groups. The distribution of CV% values from technical and biological replicates revealed that the CV% values were larger in littermate and non-

littermate groups than in the technical replicates. Analysing the variance values (CV%) at the distribution level, the results suggest that technical variance predominates in the total variance. Comparing the presented results to other similar studies (post-electrophoresis staining), a smaller technical variation can be seen. This may be due to the software analysis ("SameSpots") approach. It was reported that this approach leads to a significant reduction in the variation compared to a conventional analysis software program.

Sample size requirements were assessed by power calculation for typical fold changes in proteomic studies when the total variances (SD) encompassed 95% of the spots in littermate and non-littermate replicates. When the significance level was set to 0.05 the power calculation indicated that to reach the target power of 0.8, four replicates were required for a 2.0-fold change and nine replicates for a 1.5-fold change. The power calculation also indicated that the used sample size (n=4) was appropriate to detect around 2.0-fold change in protein expression differences. Thus, in the subsequent pairwise comparisons the 2.0-fold or higher changes were used as one of the significance criteria, although the number of spots was also calculated which presented at least 1.5-fold change difference in their expression level.

The majority of the analysed spots had larger CV% values in littermate and non-littermate groups than in technical replicates (76% of littermate spots and 67% of non-littermate spots). This also meant that around 25-30% of the spots had a negligible biological variation compared to the technical variation. The ratio of the biological variation compared to the total variation was determined by plotting the CV% values against each other. Analysing the spots which had a considerable (i.e. at least twice) biological variation compared to the technical variation, only a small fraction of spots could be found (18% of littermate spots and 17% of non-littermate spots). These results also suggest that the technical variance has a big influence on the total variance.

To avoid losing any potential protein expression differences, pairwise comparisons were made separately from the previously described analysis, which included all gels. Inter-experiment studies were performed by making pairwise comparisons using either littermate or non-littermate NMRI mouse groups. No significant protein expression differences were found in the littermate and non-littermate groups analysing the data with either FDR or SGoF multiple test statistical approaches and taking into account the magnitude of minimal fold changes. Two generations of a littermate group, and two groups of non-littermate mice were compared separately in order to detect any potential protein expression differences between consecutive mouse generations. It was assumed that more protein expression differences

could be detected if we compared two groups of mice that had about two-month difference between their birth dates. Taking into account the magnitude of the relative fold change (1.5- or 2-fold change) and the significance criteria (FDR-corrected *p*-value ≤ 0.05 and power ≥ 0.8), our analysis revealed that less than 1% of the matched spots were significantly different. However, after SGoF analysis more, significantly different protein spots could be detected. Analysing the relative fold changes, the data revealed that no or only a few spots displayed 2-fold or higher differences in their expression level.

A pairwise comparison was also performed using one of the non-littermate NMRI groups and one group of non-littermate C3H/HEN mice. NMRI is an out-bred mouse stock, while C3H/HEN is an inbred mouse strain. Our analysis also found a very similar and a quite low total variance, but the highest number of significant protein expression differences was observed in this comparison.

Proteomic study of the effect of oligomeric Aβ1-42 on SH-SY5Y cell line

Several proteomics studies have already been published in AD investigations. Despite the accumulated knowledge, the exact mechanism of Aβ1-42 induced neurodegeneration remains unclear. Recently, investigations of oligomeric Aβ induced neurodegeneration have come to the fore, as it is now widely accepted that Aβ toxicity is positively linked to Aβ oligomers.

Cell lines are good and useful models in proteomics for investigating Aβ1-42 induced neurodegeneration. In the presented study, differentiated SH-SY5Y human neuroblastoma cells were treated with oligomeric Aβ1-42 *in situ* prepared from "iso-Aβ1-42" peptide for 8h. After treatment, cell lysates were separated by 2-DE. Taking into account the significance criteria ($p \leq 0.05$ and power ≥ 0.8) and 2-fold change, 14 spots could be detected (8 down- and 6 up-regulated). To avoid loosing any important protein expression changes occurred during the treatment, the minimal fold change criterion was expanded to 1.5-fold change, where the original significance criteria were also true. According to these significance criteria, a software-based analysis revealed 52 significantly altered spots. Multiple test statistical approaches were also used to calculate the FDR- and SGoF-adjusted *p*-values. The calculated data reveal that only one spot is significantly different after FDR adjustment of *p*-values. However after SGoF analysis, all chosen spots can be accepted as significant difference

between untreated and treated groups. In addition, to verify the 2-DE results two of the most strikingly altered proteins (Hsp70 and EEF2) were confirmed by Western blot.

After a mass spectrometric analysis, 47 proteins were identified (22 down- and 25 up-regulated), as in some cases the same protein was identified from different spots. For example, the eukaryotic elongation factor-2 protein (EEF2) was observed in three adjacent spots. Each spot displayed a significant difference and the relative fold changes ranged from -2.3 to -3.6. Interestingly, heterogenous nuclear ribonucleoprotein K (hnRNP K) could be identified from two distant located spots with quite a big difference in their molecular weight. In addition, the higher molecular weight isoform of hnRNP K showed a 2.5-fold change decrease and the lower molecular weight isoform showed a 2.1-fold change increase at the expression level, which were clearly visible on the gel images. The protein identification of hnRNP K protein had a high probability and in both spots the protein was identified with a lot of unique peptides. Also, it should be mentioned that in the lower molecular weight spot only N-terminal peptides could be identified, which may account for the molecular weight difference.

The classification of the identified down- and up-regulated proteins gave well-defined functional groups. In the down-regulated protein group, six subgroups were found. Most proteins play a role in protein biosynthesis (28%), metabolic processes (22%), cytoskeleton organisation (18%) and regulation of transcription/mRNA processing (18%). The up-regulated protein group could also be divided into six subgroups, but these groups differed from the down-regulated ones. Most proteins are related to stress response (36%), protein processes like proteolysis (24%), metabolic processes (12%) and regulation of transcription/mRNA processing (12%). On the whole, it was clear from the proteins identified that the Aβ1-42 influenced some of the main cellular functioning, such as protein biosynthesis, metabolic processes, cytoskeleton organisation and transcription. The high number of up-regulated stress proteins also reflects the cellular stress triggered by Aβ1-42. What is more, most of the stress proteins have a chaperone molecular activity (seven proteins) and most of them belong to the ER chaperones and play role in ER stress.

Summary

First, the small values for the technical variation and total variation obtained from the presented 2-DE experiments (mouse brain and SH-SY5Y cell line proteome studies) indicate

that the parameters of the 2-DE workflow are well standardised. The calculated variability results were favourable when compared with previously published studies. “Samespots” analysis would accountable for smaller technical variation.

Second, analysing NMRI mice brain proteome, the proportion of biological and technical variation in total variation was estimated. These results indicate that the technical variance predominates in total variation. As these results characterise the anticipated variation using mouse brain for proteomic study hence they should be useful for future experimental design in other proteomics laboratories

Third, the effect of the genetic background on the total variation of mouse brain proteome was investigated. For this purpose I compared the variations of brain proteome profiles of littermate and non-littermate NMRI mice groups. The results displayed only a small difference. It suggests that genetic background has a negligible effect on the brain proteome profile when NMRI mouse stock is used. A comparison of the total variance of an out-bred mouse stock (NMRI) and an inbred mouse strain (C3H/HEN) also reveals a small genetic influence on the total variance, as the calculated total variance values are very similar in both populations. However, the power calculation indicates that a higher sample size would be appropriate to detect smaller differences. In addition the detection limit of the 2-DE should be taken into account. It can be assumed, that the analysed 446 protein spots belong to the high-abundant proteins. Thus a different variation could be observed analysing other subset of proteome, such as low-abundant proteins. It can be also assumed that the biological variation of other organs or other model systems may differ from the mouse brain proteome. Anyway, these results suggest that the random selection of animals as generally used in animal experimental design is feasible for neuroproteomic studies as well.

Fourth, pairwise comparisons suggest that using mice which are born in a short period of time is recommended, as no significant protein expression differences were found from analysing these groups of mice. These results reveal that the 2-DE workflow has a good reproducibility. The results also suggest that the gel-to-gel variation can be reduced by processing the samples in a short period of time. My 2-DE studies also demonstrate that a well standardized 2-DE workflow followed by a much cheaper post-stain (RuBPs) could be useable to detect higher

protein expression changes (e.g. 1.5- or 2-fold change), especially using animal models or cell lines with a relatively low biological variation.

Fifth, deeper statistical analysis reveals the importance of using different statistical approaches as different results can be obtained depending on the statistical tools. Controlling not only the type-I, but the type-II error rate and applying multiple test approaches (FDR, SGoF) are fully recommended. These statistical tools may help to avoid false discoveries, missing true differences and improve the statistical validity of the results got from a 2-DE study. However, the accepted significance cut-off values greatly depend on the user's choice. The difference arose from statistical approaches could be observed analysing either mouse brain or SH-SY5Y cell lysate.

Sixth, using quantitative 2-DE, I studied the effect of oligomeric Aβ1-42 *in situ* prepared from the precursor peptide "iso-Aβ1-42". In this experiment, differentiated SH-SY5Y neuroblastoma cells were treated with Aβ1-42. Numerous proteins displayed significantly changed expression. Moreover, two of the most striking differences (Hsp70 and EEF2) were validated by Western blot analysis. This may confirm the results got from the 2-DE analysis.

Seventh, the functional classification shows that most of the identified proteins play a role in TCA, energy production, cytoskeleton organisation, protein biosynthesis and the relatively high number of stress-related proteins reveals dramatic cellular stress induced by oligomeric Aβ1-42. Most of these cellular processes have already been related to Aβ1-42-mediated cellular stress and subsequent cell death. The strong display of ER chaperones may indicate that Aβ1-42 induced ER stress. It is well known that ER plays a crucial role in cellular Ca^{2+} homeostasis and proper protein folding process. In addition, a significant portion of the down-regulated proteins play a role in protein biosynthetic processes. It is in accordance with the ER stress. Serious or prolonged ER stress could induce cell death, thus further analysis of the related molecular pathways may help to understand the mechanism of cell death induced by Aβ1-42.

Acknowledgements

I am grateful to Prof. Botond Penke for allowing me to join his research group and for giving an opportunity to me to perform my experiments.

I would like express my gratitude to Dr. Tamás Janáky for scientific guidance, encouragement and support during my Ph.D. course.

I would like to thank everyone who helped me to perform these experiments and provided useful technical support: Dr. Gábor Juhász, Dr. Datki L. Zsolt, Dr. Bozsó Zsolt, Dr. Zoltán Szabó, József Klem, Rita Szendrei, Dóra Simon, Szilvia Pataki, Emese Mózes and Anikó M. Tóth.

I would like to thank all members of the Alzheimer's Disease Research Group and the Department of Medical Chemistry.

Finally, I am especially thankful to Adri and my family for their love and encouragement.

The research was supported by the following grants: OTKA NK 73672 from the Hungarian National Sciences Foundation; Memoload FP-7 201159; Lipididiet FP-7 211696 and TAMOP 4.2.2 project.