USING MASS SPECTROMETRY TO INVESTIGATE PROTEIN BIOMARKERS

KELLY WORMWOOD[1,2], ARMAND G. NGOUNOU WETIE[1,2], KATHERINE BEGLINGER[1], JARROD MATTINGLY[1], COLEMAN LARLEE[1], ALISA G. WOODS[1], COSTEL C. DARIE[1,*]

Keywords: mass spectrometry, proteomics, autism spectrum disorder

Abstract:
The use of mass spectrometry (MS) in the proteomics field is ever-increasing. MS provides the ability to determine minimal protein differences across numerous samples in a short amount of time. It also allows for the ability to determine what proteins are present and in what abundance. This information is very useful and applicable in many fields. MS can be used in industry as a quality control tool to determine the potential effectiveness of products, as well as in the medical and biochemical fields to determine biomarkers for disorders and diseases. Here we will discuss the importance of MS in the proteomics field and its applications in biomarker discovery, specifically in Autism Spectrum Disorder (ASD), with focus on methodology and data analysis. The importance of data processing using specialized software such as Scaffold is also demonstrated.

INTRODUCTION

The use of proteins and proteomic in biomarker identification or discovery is still in the early stages of development when compared to other techniques, such as genomics (*1-3*). A PubMed search performed on July 31, 2013 comparing genomic biomarkers to proteomic biomarkers shows over an 8-fold difference. The search also found that there are about 2 times as many studies found involving MS and proteomics compared to MS and genomics. Therefore, more information is available in this filed from genomics as compared to proteomics.

MS is a very useful method, because it is able to determine comprehensive information about proteins, such as mass, sequence information, protein post-translational modifications and protein-protein interactions (*4-9*). Depending on the specific method of MS used, it is possible to obtain protein information for thousands of proteins in a very small amount of time. This is particularly useful in biomarker discovery (*10, 11*).

All mass spectrometers contain an ion source, mass analyzer and detector. There are two types of mass spectrometers based on the ion source: Matrix-Assisted Laser Desorption Ionization mass spectrometer (MALDI-MS) and electrospray ionization mass spectrometer (ESI-MS). MS (usually ESI-MS) can also be coupled with liquid chromatography (LC) into liquid chromatography-tandem mass spectrometry (LC-MS/MS). All MS methods follow the same general format for the analysis of the proteins. The protein or peptide sample is ionized in the ion source and then it passes through the mass analyzer where the proteins/peptides are separated by their mass to charge (m/z) ratio. Finally, the peaks that correspond to these peptides/proteins reach a detector, which produces the mass spectra for the protein/peptides sample analyzed. Data processing and database search leads to identification of a protein or mixture of proteins (*4, 7, 8*).

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder with few well-understood causes. ASD has a high incidence. It is estimated that about 1 in 88 US children, 1 in 54 boys have ASD (*12*). Some other studies reported an even lower incidence of at 1:50 (*13*). Very often, ASD in children is undetected (*14, 15*). When it is detected, due to the currently used and available screening instruments (behavioral measures) it generates many false positive identifications (*16*). However, ASD treatment is effective when children are detected early in their life (*17-19*). Therefore, early detection of ASDs is critical for effective treatment. Some studies have already been performed in ASD for identification of serum or saliva biomarkers (*20-24*), however, ASD overall is very under-investigated using proteomics approaches. Here, we performed a pilot study for ASD samples and investigated the differences between the salivary proteomes of children with ASD and their matched controls. Our results show proof-of-concept that MS can easily be employed for the investigation of ASD and potentially for other neurodevelopmental disorders.

MATERIAL AND METHODS

Sera: Sera (7 with ASD and 7 controls) were provided by Taurines and colleagues; description of the sera, collected under IRB and informed consent is also provided in this paper (*22*).

Tricine-SDS-PAGE. The gels were made in the laboratory (16% acrylamide-bisacrylamide), Tricine-PAGE and were loaded with equal amounts of sera and run and then stained by Commassie according to published procedures (*25*).

Protein digestion and peptide extraction. The gels were cut in 14 gel pieces from each gel lane. Each gel piece had from every serum had a correspondent in the sera from all 7 ASD and 7 control children. The gel pieces were then digested by trypsin according to published protocols (*26*). Prior digestion, the cysteine residues from the proteins were reduced by dithiothreitol and alkylated by iodoacetamide (*26*). The resulting peptides were extracted and then combined and dried, and then solubilized in 20 μL of 0.1% FA/2% ACN in HPLC water, placed in UPLC vials and further used for LC-MS/MS analysis.

LC-MS/MS. The peptide mixtures were analyzed by reversed phase liquid chromatography and MS (LC-MS/MS) using a NanoAcuity UPLC (Micromass/Waters, Milford, MA) coupled to a Q-TOF Micro MS (Micromass/Waters, Milford, MA). The entire procedure used was previously described (*27, 28*). Calibration was performed for both precursor and product ions using 1 pmol GluFib (Glu1-Fibrinopeptide B) standard peptide with the sequence EGVNDNEEGFFSAR and the monoisotopic doubly-charged peak with m/z of 785.84.

Data processing and protein identification. The raw data were processed using ProteinLynx Global Server (PLGS, version 2.4) software as previously described (*27, 28*). The resulting pkl files were submitted for database search and protein identification to the public Mascot database search (www.matrixscience.com, Matrix Science, London, UK) using the following parameters: human databases from NCBI and SwissProt (SwissProt_2013_08 database, selected for Homo sapiens, unknown version, 20267 entries), parent mass error of 1.3 Da, product ion error of 0.8 Da, enzyme used: trypsin, one missed cleavage, and carbamidomethyl-Cysteine as fixed modification and Methionine oxidized methionine and phosphorylation of Serine, Threonine and Tyrosine as variable modification. Additional database searches were performed using the in-house PLGS database version 2.4 (www.waters.com). The Mascot and PLGS database search provided a list of proteins for each gel band. The Mascot results were exported as .dat files and then combined for ASD and for controls and uploaded on in-house Scaffold software version 4.0 (www.proteomesoftware.com).

Criteria for protein identification. Scaffold (version Scaffold_4.0.4, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 20.0% probability by the Scaffold Local false discovery rate (FDR) algorithm. Protein identifications were accepted if they could be established at greater than 99.9% probability and contained at least 1 identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm (*29*). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins were annotated with GO terms from NCBI, downloaded Sep 14, 2013 (*30*).

RESULTS AND DISCUSSIONS

Tricine-PAGE analysis of the sera from children with ASD and their matched controls. Several years ago, Taurines et al (*22*) identified a biomarker signature, with a specific peak with a mass-to-charge ratio (m/z) of almost 11 kDa. However, the method used (MALDI-MS) did not provide sequence information. Therefore, our goal was to identify the protein(s) responsible for the 11 kDa peak and to perform a full proteomics analysis of the sera from children with ASD and their matched controls. As such, our overall strategy involved fractionation of the sera under Tricine-PAGE under non-reducing conditions, followed by enzymatic digestion and LC-MS/MS analysis and data processing. Tricine-PAGE separates proteins under denaturing conditions, just like SDS-PAGE, except that the resolution of Tricine-PAGE is much higher for the low molecular mass proteins and peptides (2-15 kDa) as compared with SDS-PAGE. In addition, we separated the sera under non-reducing conditions, because it allowed us to keep proteins such as Haptoglobin and Immunoglobulin G (IgG) in their oxidized, disulfide-linked form, thus preventing their interference with our analysis. Therefore, none of the abundant proteins interfered with our experiments.

LC-MS/MS analysis of the sera from children with ASD and their matched controls. For LC-MS/MS analysis, we cut out the gel bands from all 14 samples (they corresponded to molecular masses between 60 and 5 kDa), digested the by trypsin and analyzed them by LC-MS/MS followed by protein identification using Mascot database search and data analysis by Scaffold software. Using Scaffold software, we identified a series of proteins, some of them shown in the print-screen shot shown in Figure 2. In this figure, information about each specific protein is shown. For example, information about the identified proteins in terms of biological process,

cellular component or molecular function, as well as the quantitative value based on the number of spectra that led to identification of that protein is also shown (Figure 2). In addition, we not only identified the proteins and the differences between the protein composition from the two sets of sera, but we also identified the amino acid coverage for each protein, within each condition (ASD or controls), protein quantitation, and protein post-translational modification (i.e. phosphorylation or acetylation; currently it indicates no filter in Figure 2). Furthermore, we were also able to verify the amino acid sequence coverage for each protein within both conditions (ASD and controls). Figure 3A & B shows such an example for Transthyretin. We were also able to obtain information within the reconstituted mass spectra for peptides of interest with the amino acid sequence GSPAINVAVHVFR which is part of Transthyretin or HLSLLTTLSNR which is part of Vitamin D binding protein (Figure 5A). We can also verify and locate the post-translation modifications (PTMs) within a particular peptide such as experimental PTMs (alkylation of iodoacetamide to carbamydomethylcysteine or methionine oxidation; Figure 4B), natural PTM (i.e. threonine phosphorylation; Figure 4C) or a combination of both (i.e. methionine oxidation and lysine acetylation; Figure 4C).

In these experiments, we also classified the proteins and peptides according to the ASD and controls (Figure 5) and their classification according to the biological process, cellular process and molecular function (Figure 6). We did not quantify the identified proteins, but we do have this option, using label-free, spectral counting that reflects the fold change in the levels of proteins in ASD and controls (followed by Fisher’s exact test or when appropriate, using the Student’s T-test). However, we did identify some proteins that are specific to ASD or to controls. For example, alpha-2-macroglobulin was specific to ASD, but not controls (20 spectra in ASD but none in the controls). Similarly, IgG kappa chain C region was specific to ASD, but not controls (19 spectra in ASD but none in the controls). An opposite effect was observed for protein Piccolo (16 spectral in controls, but only 2 in the ASD) or for PHD finger protein 12 (10 spectra in controls but none in ASD). Although these differences sound very promising, careful interpretation of the results and inspection of the raw data still must be performed. However, these results should be carefully inspected. For example, while the quality of the spectra that led to identification of alpha-2-macroglobulin and IgG kappa chain C region is acceptable, the quality the spectra that led to identification of PHD finger protein 12 and Piccolo are not high enough. Therefore, when in doubt, additional inspection of the raw data should be performed.

Other factors that should be considered when investigating proteins at the proteome scale include, among others, sample manipulation (i.e. biochemical fractionation). For example, we separated sera under denaturing and non-reducing conditions. The proteins investigated were in the 2-60 kDa range. Yet, three of the four proteins mentioned above had a molecular mass much higher than the separation range (2-60 kDa). For example, the mass of Alpha-2-macroglobulin is 162 kDa, of PHD finger protein 12 is 110 kDa and of Protein Piccolo is 553 kDa, while only the mass of IgG kappa chain C region was in the investigated molecular mass range (12 kDa). Therefore, although the automated software is important and extremely helpful, additional verification of the investigated proteins is also required.

CONCLUSIONS

Our results show proof-of-concept that MS can easily be employed for investigation of ASD and of other disorders and diseases for protein identification, protein characterization, PTMs, protein-protein interactions and protein quantitation. In these preliminary studies, we did not identify the

protein(s) responsible for the 11 kDa peak, as previously published. However, this does not mean that the peak is not there. It is, just that we have to use alternative approaches to identify it (them), currently under investigation. Overall, we also identified some differences at the protein pattern in the sera of the children with ASD and matched controls, which are currently being investigated. This is a starting platform for a comprehensive proteomic investigation of the ASD and matched control samples and for additional proteomics research.

REFERENCES

1. Darie, C. C. (2013) Mass Spectrometry and Proteomics: Principle, Workflow, Challenges and Perspectives, *Mod Chem appl 1*, e105.
2. Darie, C. C. (2013) Investigation of Protein-Protein Interactions by Blue Native-PAGE & Mass Spectrometry, *Mod Chem appl 1*, e111.
3. Darie, C. C. (2013) Mass spectrometry and its application in life sciences, *Australian Journal of Chemistry 66*, 1-2.
4. Ngounou Wetie, A. G., Sokolowska, I., Woods, A. G., and Darie, C. C. (2013) Identification of post-translational modifications by mass spectrometry, *Australian Journal of Chemistry*.
5. Ngounou Wetie, A. G., Sokolowska, I., Woods, A. G., Roy, U., Deinhardt, K., and Darie, C. C. (2013) Protein-protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches, *Cellular and molecular life sciences : CMLS*.
6. Ngounou Wetie, A. G., Sokolowska, I., Woods, A. G., Roy, U., Loo, J. A., and Darie, C. C. (2013) Investigation of stable and transient protein-protein interactions: Past, present, and future, *Proteomics 13*, 538-557.
7. Sokolowska, I., Ngounou Wetie, A. G., Woods, A. G., and Darie, C. C. (2013) Applications of mass spectrometry in proteomics, *Australian Journal of Chemistry*.
8. Sokolowska, I., Woods, A. G., Wagner, J., Dorler, J., Wormwood, K., Thome, J., and Darie, C. C. (2011) Mass spectrometry for proteomics-based investigation of oxidative stress and heat shock proteins, In *Oxidative Stress: Diagnostics, Prevention, and Therapy* (Andreescu, S., and Hepel, M., Eds.), American Chemical Society, Washington, D.C.
9. Woods, A. G., Sokolowska, I., Yakubu, R., Butkiewicz, M., LaFleur, M., Talbot, C., and Darie, C. C. (2011) Blue native page and mass spectrometry as an approach for the investigation of stable and transient protein-protein interactions, In *Oxidative Stress: Diagnostics, Prevention, and Therapy* (Andreescu, S., and Hepel, M., Eds.), American Chemical Society, Washington, D.C.
10. Thome, J., Coogan, A. N., Woods, A. G., Darie, C. C., and Hassler, F. (2011) CLOCK Genes and Circadian Rhythmicity in Alzheimer Disease, *Journal of aging research 2011*, 383091.
11. Woods, A. G., Sokolowska, I., Taurines, R., Gerlach, M., Dudley, E., Thome, J., and Darie, C. C. (2012) Potential biomarkers in psychiatry: focus on the cholesterol system, *Journal of cellular and molecular medicine 16*, 1184-1195.
12. Autism, Developmental Disabilities Monitoring Network Surveillance Year Principal, I., Centers for Disease, C., and Prevention. (2012) Prevalence of autism spectrum disorders--Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008, *Morbidity and mortality weekly report. Surveillance summaries 61*, 1-19.
13. Blumberg, S. J., Bramlett, M. D., Kogan, M. D., Schieve, M. A., Jones, J. R., and Lu, M. C. (2013) Changes in Prevalence of Parent-reported Autism Spectrum Disorder in School-aged U.S. Children: 2007 to 2011–2012, *National Health Statistics Reports 65*, 1-12.
14. Dawson, G. (2010) Recent advances in research on early detection, causes, biology, and treatment of autism spectrum disorders, *Current opinion in neurology 23*, 95-96.
15. Zwaigenbaum, L. (2010) Advances in the early detection of autism, *Current opinion in neurology 23*, 97-102.
16. Dereu, M., Roeyers, H., Raymaekers, R., Meirsschaut, M., and Warreyn, P. (2012) How useful are screening instruments for toddlers to predict outcome at age 4? General development, language skills, and symptom severity in children with a false positive screen for autism spectrum disorder, *European child & adolescent psychiatry 21*, 541-551.
17. Eikeseth, S., Klintwall, L., Jahr, E., and Karlsson, P. (2012) Outcome for children with autism receiving early and intensive behavioral intervention in mainstream preschool and kindergarten settings, *Res Autism Spec Dis 6*, 829-835.

18. Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., and Cross, S. (2009) Meta-analysis of Early Intensive Behavioral Intervention for children with autism, *Journal of clinical child and adolescent psychology : the official journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53 38*, 439-450.
19. Lovaas, O. I. (1987) Behavioral treatment and normal educational and intellectual functioning in young autistic children, *Journal of consulting and clinical psychology 55*, 3-9.
20. Momeni, N., Bergquist, J., Brudin, L., Behnia, F., Sivberg, B., Joghataei, M. T., and Persson, B. L. (2012) A novel blood-based biomarker for detection of autism spectrum disorders, *Translational psychiatry 2*, e91.
21. Momeni, N., Brudin, L., Behnia, F., Nordstrom, B., Yosefi-Oudarji, A., Sivberg, B., Joghataei, M. T., and Persson, B. L. (2012) High complement factor I activity in the plasma of children with autism spectrum disorders, *Autism research and treatment 2012*, 868576.
22. Taurines, R., Dudley, E., Grassl, J., Warnke, A., Gerlach, M., Coogan, A. N., and Thome, J. (2011) Proteomic research in psychiatry, *Journal of psychopharmacology 25*, 151-196.
23. Castagnola, M., Messana, I., Inzitari, R., Fanali, C., Cabras, T., Morelli, A., Pecoraro, A. M., Neri, G., Torrioli, M. G., and Gurrieri, F. (2008) Hypo-phosphorylation of salivary peptidome as a clue to the molecular pathogenesis of autism spectrum disorders., *J Proteome Res 7*, 5327-5332.
24. Corbett, B. A., Kantor, A. B., Schulman, H., Walker, W. L., Lit, L., Ashwood, P., Rocke, D. M., and Sharp, F. R. (2007) A proteomic study of serum from children with autism showing differential expression of apolipoproteins and complement proteins, *Molecular psychiatry 12*, 292-306.
25. Darie, C. C., Biniossek, M. L., Gawinowicz, M. A., Milgrom, Y., Thumfart, J. O., Jovine, L., Litscher, E. S., and Wassarman, P. M. (2005) Mass spectrometric evidence that proteolytic processing of rainbow trout egg vitelline envelope proteins takes place on the egg, *The Journal of biological chemistry 280*, 37585-37598.
26. Darie, C. C., Deinhardt, K., Zhang, G., Cardasis, H. S., Chao, M. V., and Neubert, T. A. (2011) Identifying transient protein-protein interactions in EphB2 signaling by blue native PAGE and mass spectrometry, *Proteomics 11*, 4514-4528.
27. Florian, P. E., Macovei, A., Lazar, C., Milac, A. L., Sokolowska, I., Darie, C. C., Evans, R. W., Roseanu, A., and Branza-Nichita, N. (2013) Characterization of the anti-HBV activity of HLP1-23, a human lactoferrin-derived peptide, *Journal of medical virology 85*, 780-788.
28. Sokolowska, I., Dorobantu, C., Woods, A. G., Macovei, A., Branza-Nichita, N., and Darie, C. C. (2012) Proteomic analysis of plasma membranes isolated from undifferentiated and differentiated HepaRG cells, *Proteome science 10*, 47.
29. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry, *Analytical chemistry 75*, 4646-4658.
30. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nature genetics 25*, 25-29.

1 - Biochemistry and Proteomics Group. Department of Chemistry and Biomolecular Science. Clarkson University
8 Clarkson Avenue, Potsdam, NY 13699-5810, USA
2 - These authors contributed equally to this manuscript
* cdarie@clarkson.edu

Acknowledgements:
This work was supported partially by the US Army Research Office (DURIP grant #W911NF-11-1-0304 to C.C.D.), private donations (Mary Stewart Joyce, Kenneth Sandler), the Alexander von Humboldt Foundation (A.G.W.) and the generosity of SciFund Challenge-3 donors. LKW and CCD were supported during the Summer 2013 by the David A. Walsh '67 Fellowship. CCD is also thankful to Prof. Dr. Vlad Artenie for his guidance early during his career and to his teachers from the Departments of Biology and Chemistry from University of Iasi.

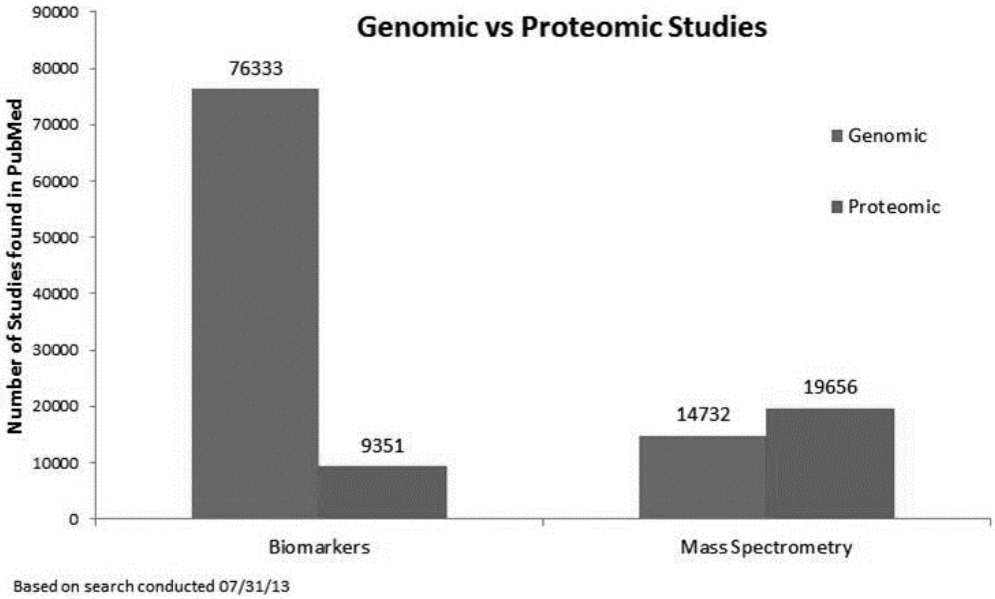

Figure 1. PubMed search of genomics versus proteomics

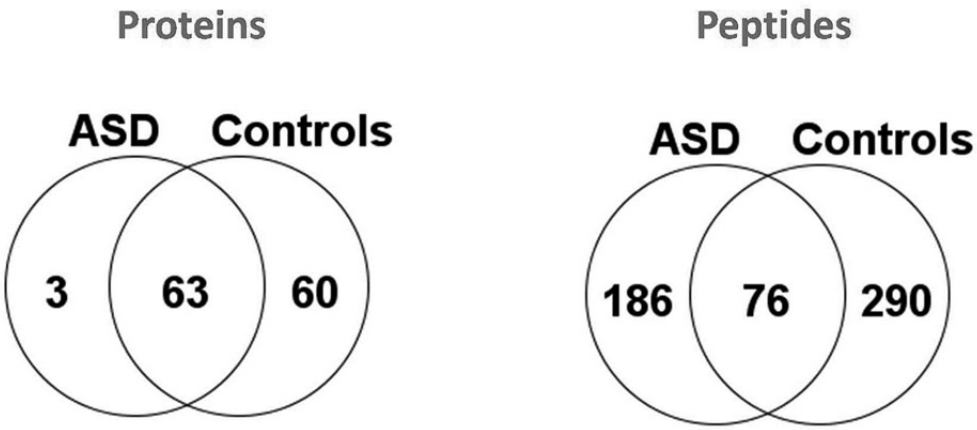

Figure 5B. Venn diagrams with all proteins and peptides identified in ASD and controls

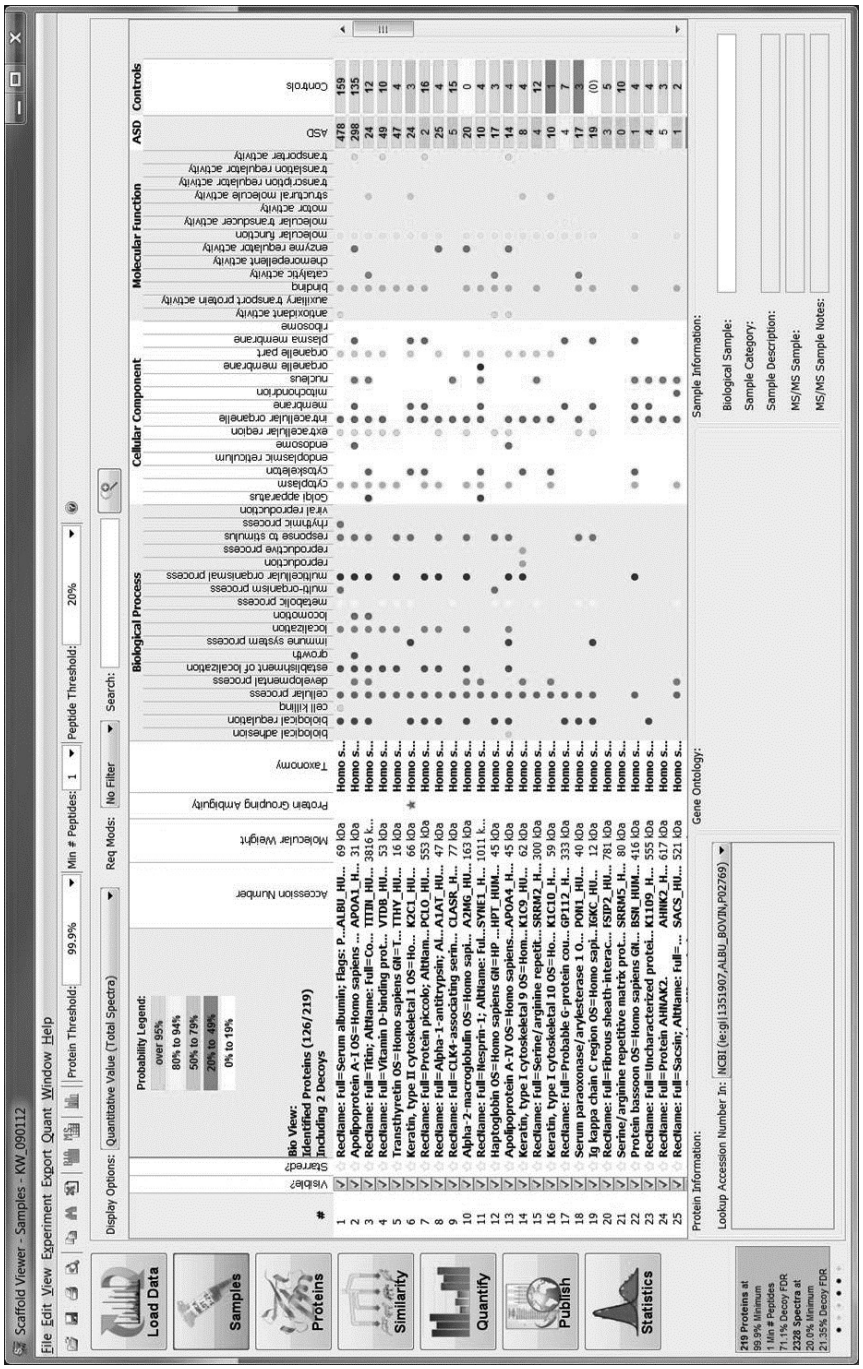

Figure 2. Overview of the Scaffold software for identification of the proteins from children with ASD and matched controls.

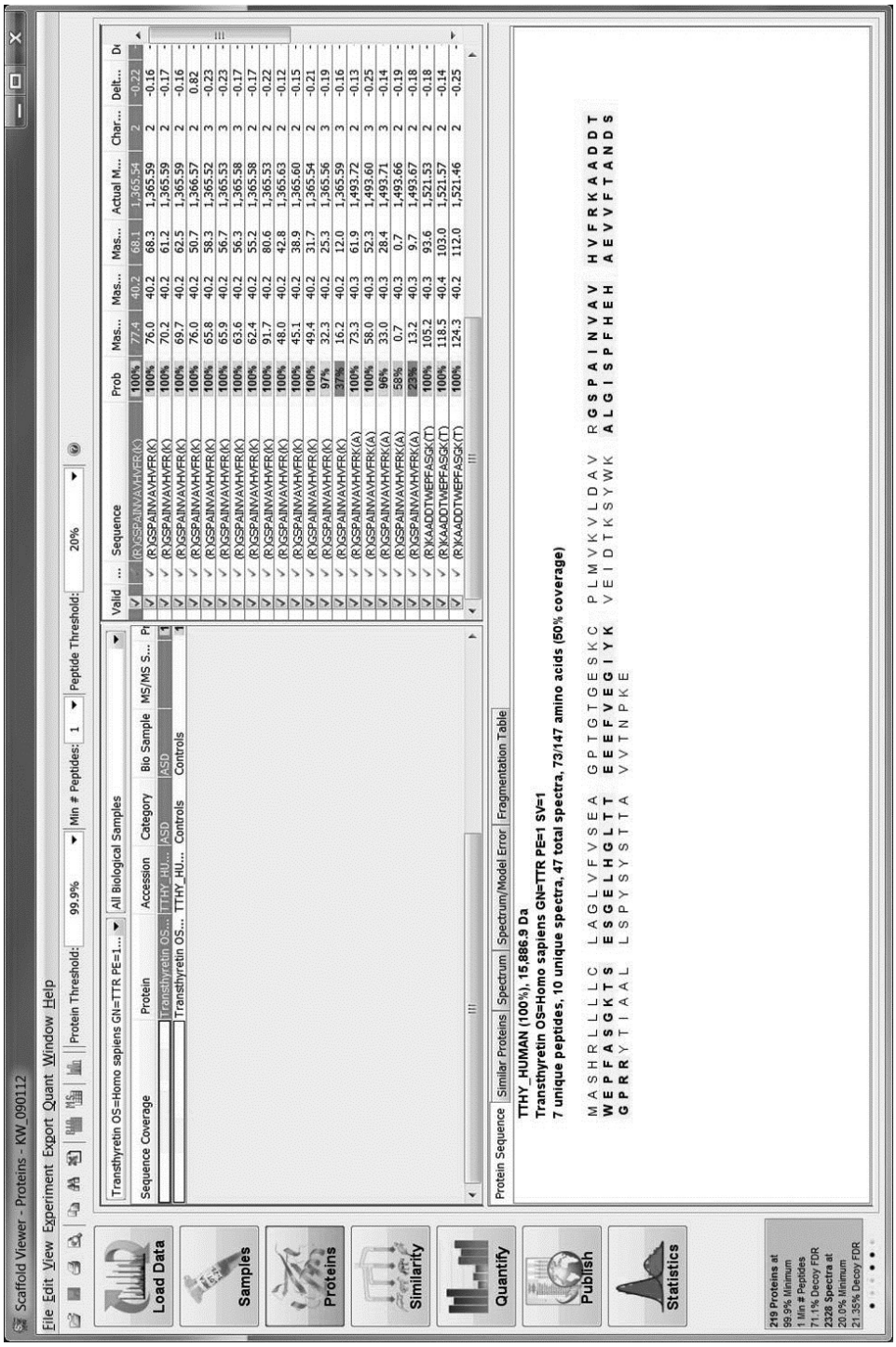

Figure 3A. Overview of the peptides that were part of Transthyretin and were identified ASD using Scaffold software

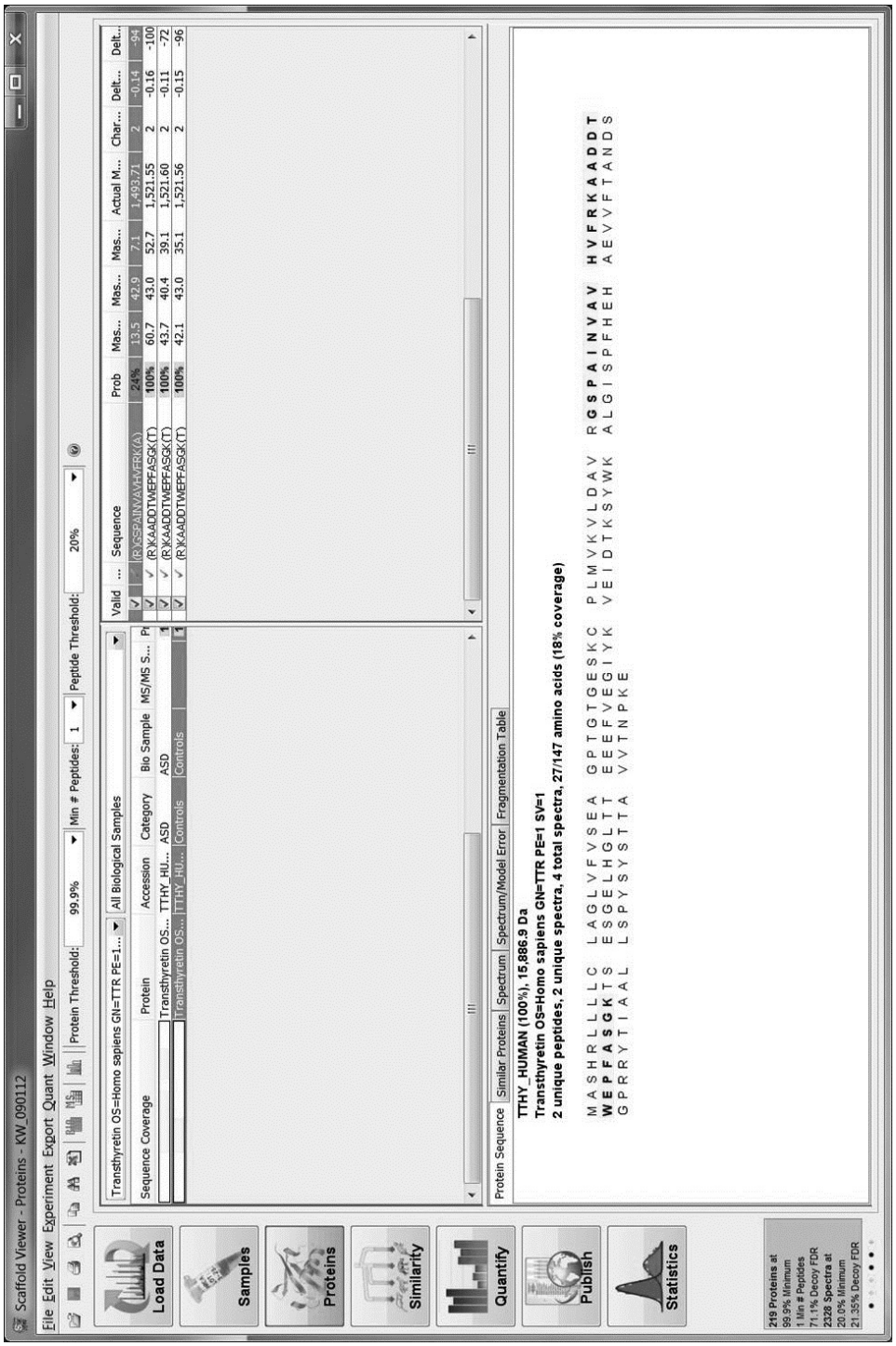

Figure 3B. Overview of the controls using Scaffold software.

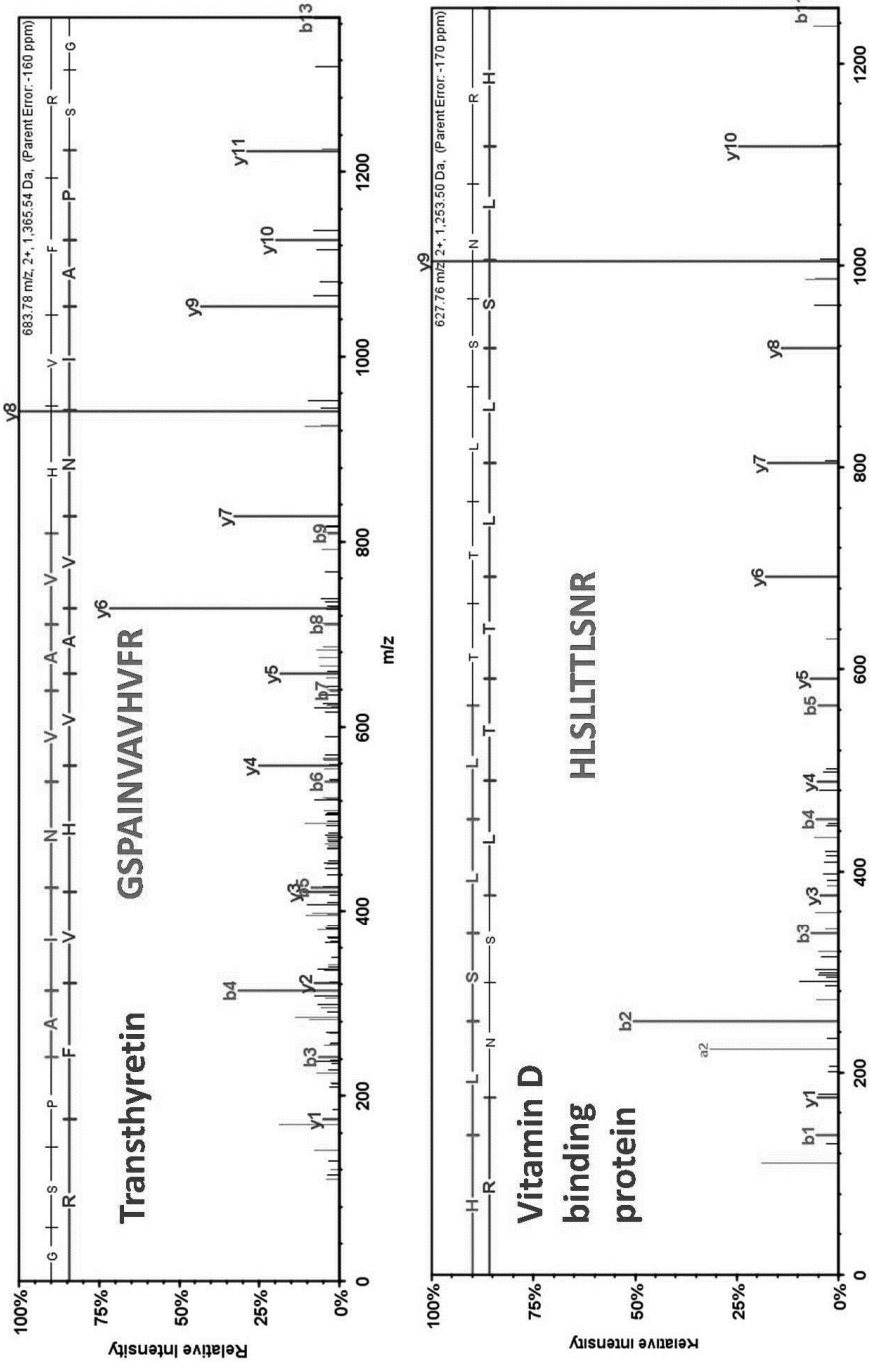

Figure 4A. Identification of spectra that correspond to peptides that were part of Transthyretin (amino acid sequence GSPAINVAVHVFR) or Vitamin binding protein (amino acid sequence HLSLLTTLSNR).

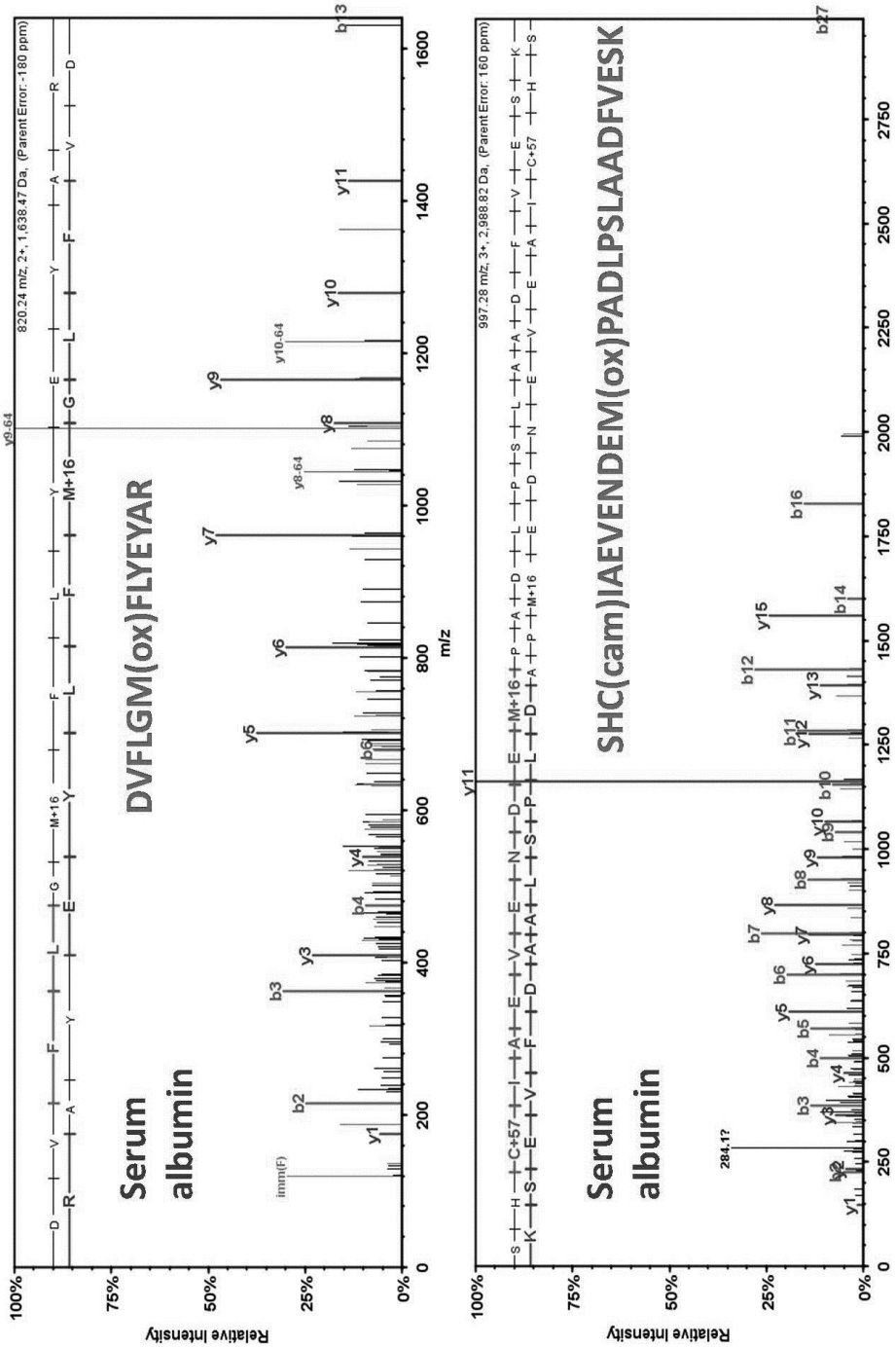

Figure 4B. Identification of peptides that were part of serum albumin with the amino acid sequences DVFLGM(ox)FLYEYAR (with methionine oxidized) and

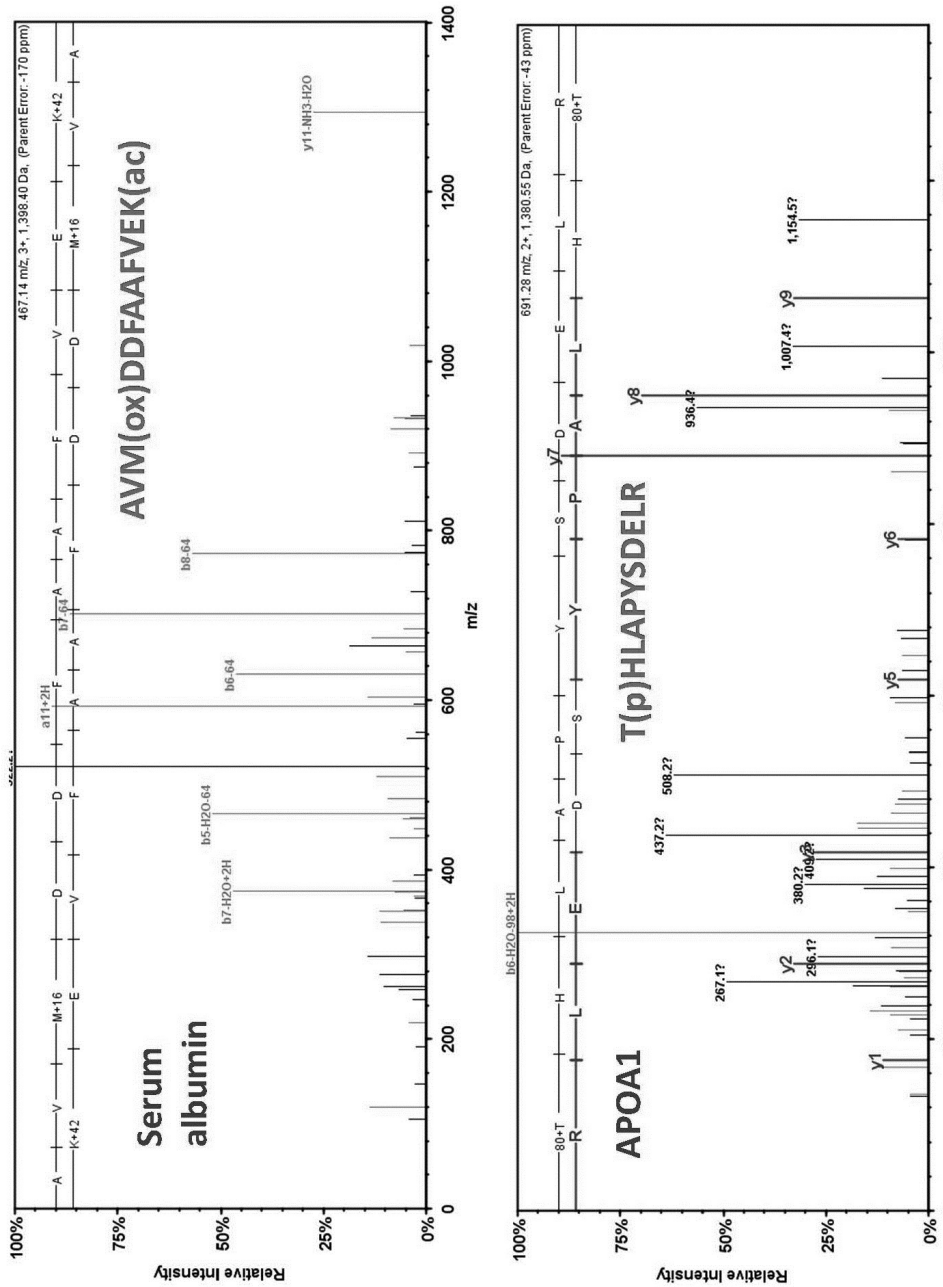

Figure 4C. Identification of peptides that were part of serum albumin with the amino acid sequences AVM(ox)DDFAAFVEK(ac) (with methionine oxidized and lysine acetylated) and T(p)HLAPYSDELR (with

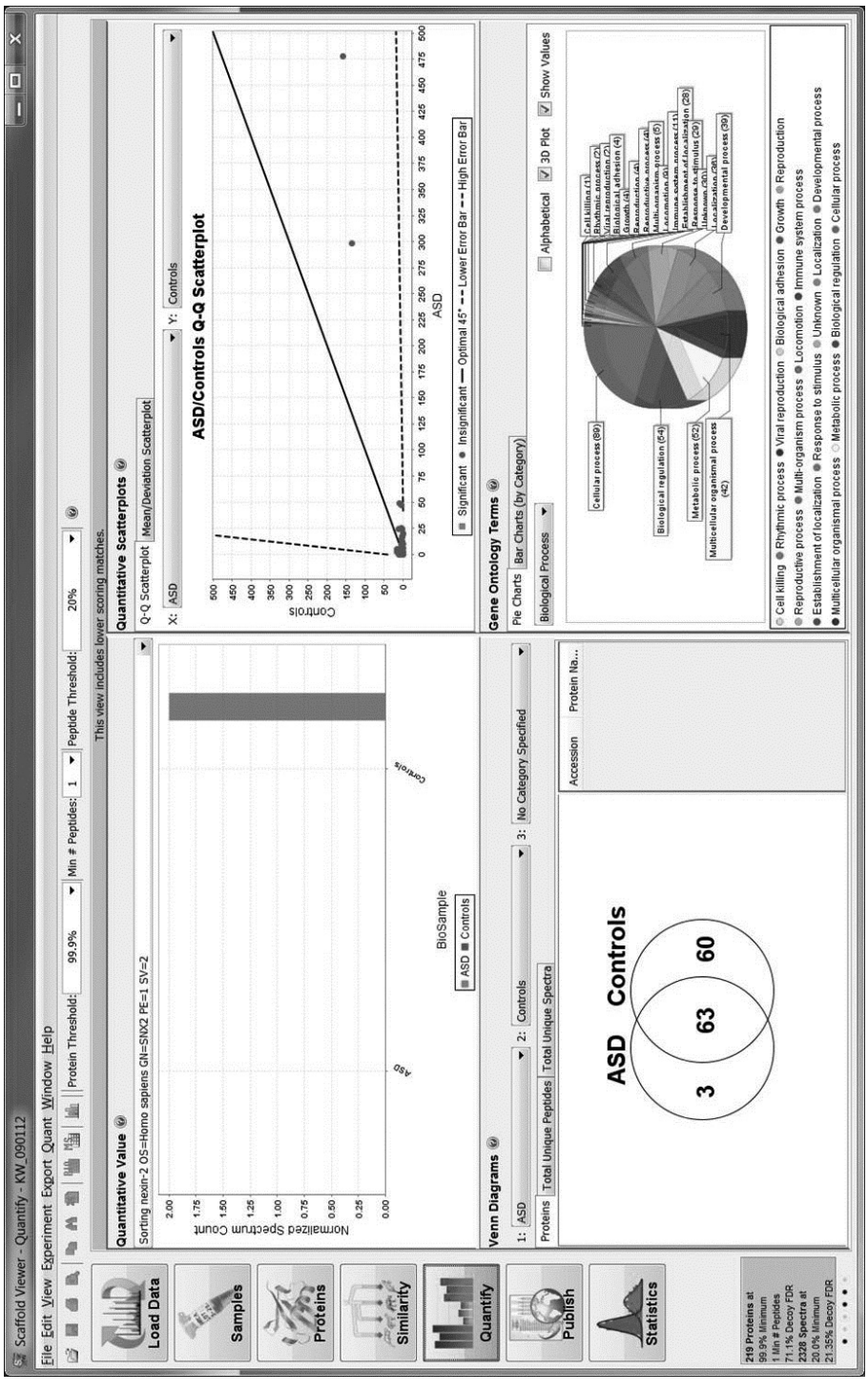

Figure 5A. Overview of the Scaffold software that shows the quantitative analysis for individual proteins (upper left), Venn diagrams with all proteins and peptides identified in ASD and controls (lower left), statistical

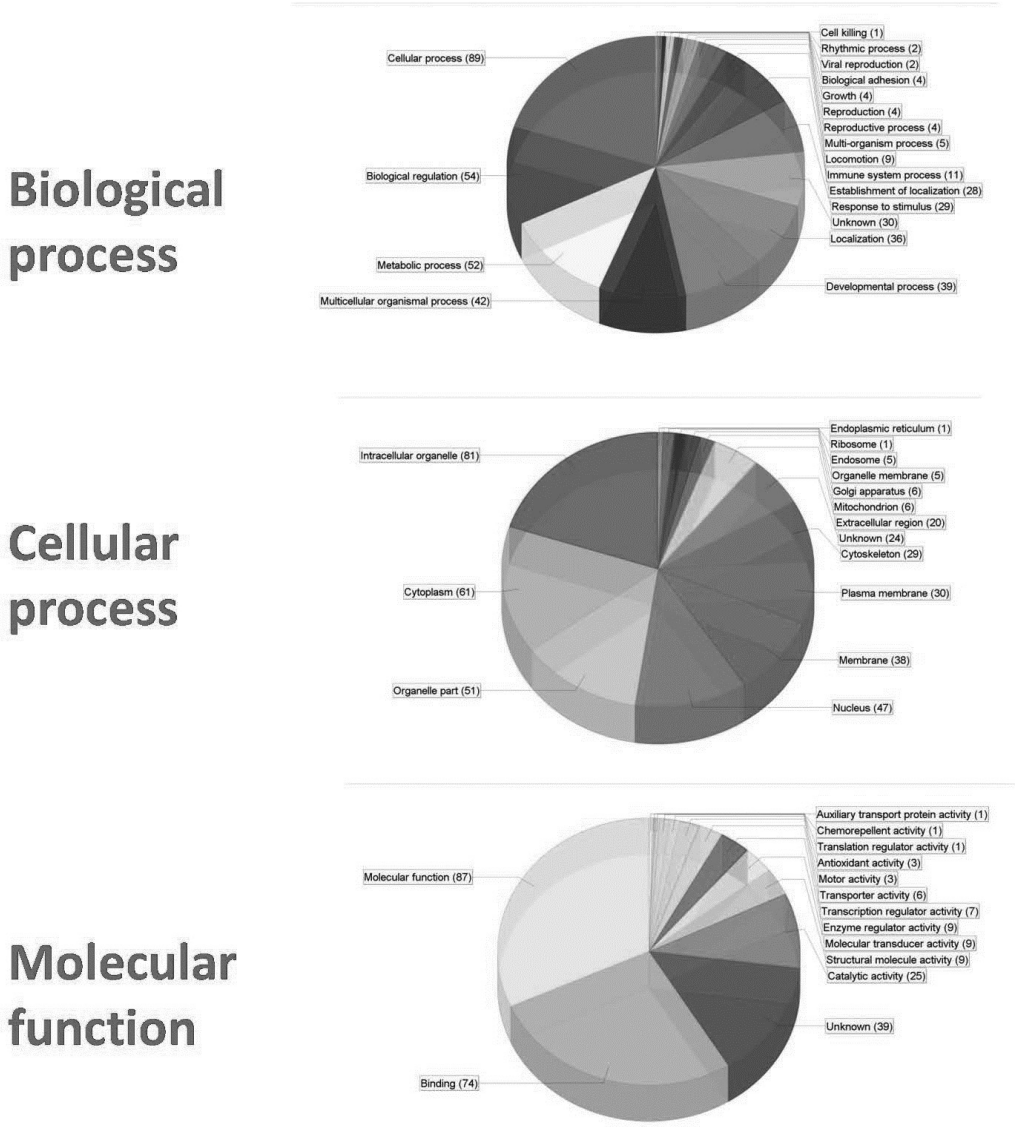

Figure 6. Classification of the proteins from ASD and controls according to the biological process, cellular process, and the molecular function.