

Visualization and Analysis of Molecular Scanner Peptide Mass Spectra

Markus Müller¹, Robin Gras¹, Willy V. Bienvenu², Denis F. Hochstrasser^{2,3}
and Ron D. Appel^{1,3}

¹*Swiss Institute of Bioinformatics, Geneva, Switzerland*

²*Clinical Chemistry Laboratory, Geneva University Hospital, Geneva, Switzerland*

³*University of Geneva, Geneva, Switzerland*
markus.mueller@isb-sib.ch

At present, as complete genomes for an increasing number of organisms are available, attention must be focused on proteins encoded by the genes. In contrast to the static genome, the proteome of an organism is a highly dynamic and connected network, and new analytical methods have to be developed in order to describe its spatial and temporal changes and interactions. An important step in this task is the automated, high throughput identification of proteins, which nowadays mostly relies on efficient protein separation, mass spectrometry, protein sequence databases as well as bioinformatics.

One of the most important methods for protein separation is 2-dimensional polyacrylamide gel electrophoresis (2D-PAGE). This technique allows separating simultaneously thousands of proteins according to their isoelectric point (pI) and molecular weight (MW) and displaying them on a two-dimensional map. Mass spectrometry (MS) has become one of the most powerful techniques to identify organic molecules. Among various applications, peptide mass fingerprinting (PMF) is frequently used because, combined with matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry, it provides a rapid and sensitive method for protein identification. PMF compares the list of experimental peptide masses, the peptide mass fingerprint, obtained by specific endoproteolytic digestion of proteins with the theoretical mass values calculated by *in silico* digestion of protein sequences. A score evaluates how well the theoretical masses match the fingerprint.

Several partially automated methods have been proposed to excise protein spots from a stained gel, to submit the excised material to endoproteolytic digestion and to extract peptides from the excised gel. The peptides are then loaded onto a MALDI sample plate and introduced into a mass spectrometer for PMF acquisition. These methods have the inconvenience that the location of protein spots must be known prior to excision, and that the excision precision is limited (> 1mm). We introduced a new and highly automated approach, dubbed the molecular scanner, which combines 2-D PAGE separation techniques with PMF methods (Binz et al., 1999). In this approach, the proteins are digested firstly in the gel itself and then during transfer onto a collecting polyvinylidene fluoride (PVDF) membrane. This membrane is sprayed with a matrix solution, and the

co-crystallisation of the matrix and the peptides allows MALDI-MS analysis. Since diffusion in this process is not relevant, the location of the peptides on the PVDF membrane corresponds to the location of their proteins in the gel. The membrane is then scanned by a MALDI-TOF mass spectrometer, which yields a peptide mass fingerprint for each scanned point.

In the molecular scanner experiment discussed here, *E.coli* proteins were separated by 2D-PAGE and a portion of the PVDF membrane corresponding to a pI range of 5.1-5.2 and a MW range of 35'000-45'000 Da was scanned on a 48x32 grid with a spacing of 0.25mm (Bienvenut et al., 1999). The first aim of the work presented here (Müller et al., submitted) was to get an idea of how the data were structured. Since there were many spectra, it was impossible to inspect and compare them by means of conventional visualization tools that are only able to render a few spectra at a time. We designed a method that allows circumventing this problem and inspecting all spectra at once. This revealed an interesting feature: some masses covered the entire membrane while others were localized in spots. The ubiquitous masses could be identified as chemical noise and were purged from the fingerprint because they impaired the identification process (Figure 1).

Masses detected over the entire membrane could be used to investigate the calibration of the mass spectrometer. This revealed that mass values were locally quite stable, but varied significantly over the entire membrane, whereas the difference between the minimal and maximal measured value was about 1 Da because the membrane was warped at its upper edge (high Mr values), and because physical conditions as electric field strength depend on the position of the sampling plate. Therefore it was impossible to assign precise mass values valuable for all spectra. A re-calibration of the spectra would facilitate data handling, and we had to devise a method that does not rely on internal standard masses since these were not used in this experiment. We chose an iterative method that started from a few master spectra that could easily be calibrated since they provided very clear identifications, using the matching peptide masses as reference masses, and then used the strong correlation between neighboring spectra to spread the calibration over the entire membrane. This method successfully reduced the calibration error to 0.1 Dalton.

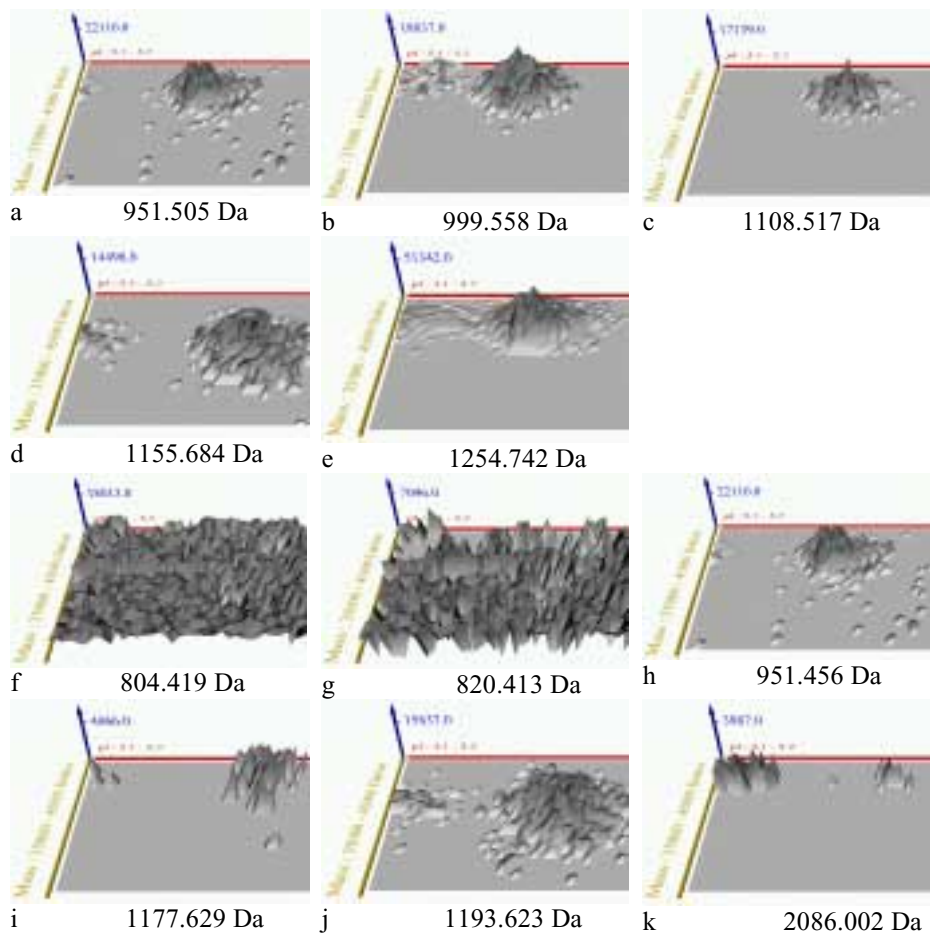


Figure 1. The vertical axis represents the peptide signal intensities (peptide signal heights) as a function of the position on the membrane. The intensity was set to 0 if no peptide mass could be detected at the respective position within ± 100 ppm of the theoretical peptide mass. Note that the scale varies from case to case. (a)-(e) Intensity distribution of the matching peptides of METK_ECOLI. (f)-(k) Intensity distribution of the matching peptides of ALLC_ECOLI. ALLC_ECOLI turned out to be a false identification.

Since the resolution of the 2D-PAGE is limited, protein spots can overlap and this enhances the possibility for false PMF identifications. Figure 1 shows an example of such a false identification in the overlapping zone of two proteins (SwissProt entries METK_ECOLI and IDH_ECOLI). A peptide mass fingerprint situated in the overlapping zone was submitted to SmartIdent (Gras

et al., 1999), which returned a list of matching proteins: IDH_ECOLI (score: 818034.11), ALLC_ECOLI (score: 51341.17), METK_ECOLI (score: 25122.18), ... Though ALLC_ECOLI had a higher score than METK_ECOLI, its matching peptides showed no consistent distribution. In fact, these were random matches with peptides from METK_ECOLI, IDH_ECOLI and chemical noise.

Therefore the fingerprints have to be untangled before identification. Peptide signal intensities as a function of their position on the membrane should positively correlate with the concentration of the protein the peptides stem from. Therefore, clustering all peptides with a similar signal intensity distribution should put all peptides originating from one protein in the same group and should separate tangled fingerprints. We calculated the intensity distribution of all 124 masses that were not part of the chemical noise and could be reproducibly detected over a part the membrane, and performed a hierarchical cluster analysis yielding a set of 20 clusters. The masses of each cluster were then submitted to SmartIdent, which found clear identifications for 13 of these 20 clusters (Figure 2). This method is fully automated and provides clear identifications also for weakly expressed proteins that would be very difficult to detect using conventional methods that perform identifications without taking account of the spatial correlation.

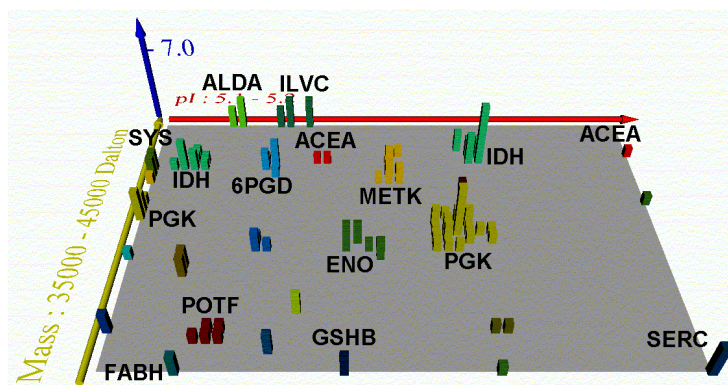


Figure 2. Summits of the intensity distributions of all 124 masses. The intensity distributions were smoothed using a median filter before the summits were calculated. The vertical axis indicates the number of summits found on the respective scan point. The groups that could be identified (13 of 20) carry a label: Aldehyde dehydrogenase A (ALDA_ECOLI), Ketol-acid reductoisomerase (ILVC_ECOLI), Seryl-tRNA synthetase (SYS_ECOLI), Isocitrate dehydrogenase (IDH_ECOLI), 6-Phosphogluconate dehydrogenase

(6PGD_ECOLI), Isocitrate lyase (ACEA_ECOLI), s-Adenosylmethionine synthetase (METK_ECOLI), Phosphoglycerate kinase (PGK_ECOLI), Enolase (ENO_ECOLI), Putrescine-binding periplasmic protein [precursor] (POTF_ECOLI), 3-Oxoacyl-[acyl-carrier-protein] synthase III (FABH_ECOLI), Glutathione synthetase (GSHB_ECOLI), Phosphoserine aminotransferase (SERC_ECOLI). IDH_ECOLI, PGK_ECOLI and ACEA_ECOLI were found on two spots.

References

Bienvenut WV, Sanchez JC, Karmime A, Rouge V, Rose K, Binz PA, Hochstrasser DF. 1999. Toward a clinical molecular scanner for proteome research: parallel protein chemical processing before and during western blot. *Anal Chem* 71:4800-4807.

Binz PA, Müller M, Walther D, Bienvenut WV, Gras R, Hoogland C, Bouchet G, Gasteiger E, Fabbretti R, Gay S, Palagi P, Wilkins MR, Rouge V, Tonella L, Paesano S, Rossellat G, Karmime A, Bairoch A, Sanchez JC, Appel RD, Hochstrasser DF. 1999. A molecular scanner to automate proteomic research and to display proteome images. *Anal Chem* 71:4981-4988.

Gras R, Müller M, Gasteiger E, Gay S, Binz PA, Bienvenut W, Hoogland C, Sanchez JC, Bairoch A, Hochstrasser DF, Appel RD. 1999. Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring algorithm and an optimized peak detection. *Electrophoresis* 20:3535-3550.

Müller M, Gras R, Bienvenut W, Hochstrasser D, Appel RD. 2001. Visualization and Analysis of Molecular Scanner Peptide Mass Spectra. *submitted*.