

Automatic acquisition of MALDI-TOF mass spectra

D. Suckau*, L. Cornett and K.O. Kräuter

Bruker Daltonik GmbH, Fahrenheitstrasse 4, D-28359 Bremen, Germany

Matrix-assisted laser desorption/ionisation mass spectrometry is an increasingly valuable tool in proteome and genome research. By definition, this kind of research involves analyses of hundreds or thousands of samples, which can only be put into practice if the huge workload is transformed into automated tasks. The principles of automated acquisition of MALDI spectra are described and example spectra discussed.

The recent development of pulsed ion extraction (PIE) [1] and routine sample preparation protocols for peptides enable high throughput analyses using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF). Especially in fields like proteome research or genetic screening, automation of the two stages of the analysis, i.e., (1) the preparation of the samples, and (2) the acquisition and processing is highly desirable in order to deal with the hundreds of samples associated with such studies.

Typically, the acquisition of MALDI spectra requires the instrument operator to carefully look at the data obtained from a series of, e.g., 10 laser shots. Then he has to decide whether to accept the spectrum because of good resolution, peak shape and intensity. If not, the acquisition conditions need to be optimised: find a better spot on the target and try again or adjust the laser fluence either to increase signal intensity or to increase the peak resolution. Also the operator has a sample type-specific experience like, e.g., “do not exceed 100 shots per spot because of “burn out” phenomena” or “first blast off the salt containing surface layer before starting acquisition at reduced laser fluence”. In the end, all these experiences and informations from the last acquired spectra result in the decision of whether to accumulate a spectrum into an average spectrum, whether to change the spot or the laser fluence.

Therefore, good operator-like performance of automated acquisition algorithms depends on the ability to transform such human experience into robust algorithms in a simple way. Fuzzy logic is an approach to solve problems by adapting a human-like perception of the world and decision-making processes, which suggested it is an adequate strategy to convert human experience into automated procedures.

Fuzzy logic

As a simple example, we implement an office radiator control, which adjusts the fan speed as a function of

temperature (Fig. 1). First, an input parameter, called “temperature” is defined, which has associated set functions like “warm” or “cold”. These functions can assume all values between 0, “false”, and 1, “true”. The term fuzzy logic actually comes from the fact that there are not just the digital logic levels 0 and 1, but also intermediate states like “60% true” (0.6) which corresponds more to human perception. At 23 °C the function “warm” may be 60% true and “fine” 40% true, while all other set functions are 0 at 23 °C. In a similar way the output parameter “radiator speed” has associated functions like “slow” or “fast”. Their values are calculated as a function of the input parameter set functions and a set of rules. A quite simple set of rules is used here, like: “If “temperature” is “fine” then “speed” is “fast””. Therefore, if “warm” is 60% true, then the related output function “fast” is 60% true. The area under the output function values is averaged with respect to the output parameter which results in a specific output value of the fuzzy logic control module (defuzzification). In this case, the radiator is set to 67% of the maximum speed.

But why using fuzzy logic for this simple control, why not just have a linear function instead? The real beauty of fuzzy logic turns out when you want to change the performance in some aspect. E.g., you want to redefine “warm” or “fine” to see whether people feel better with a radiator controlled that way, or you add the input parameter

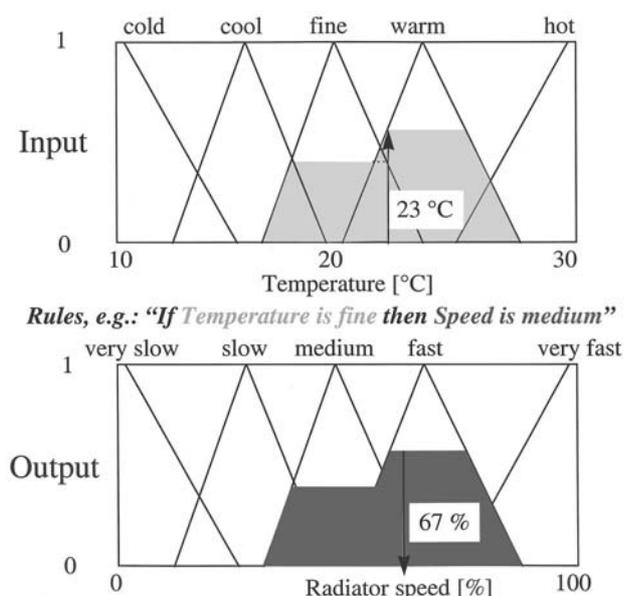


Figure 1. Example for a fuzzy logic control: the speed of an office radiator is defined by the fuzzy set functions which define the input parameter “temperature” and a set of fuzzy rules.

“humidity”. Such changes can be implemented in a fuzzy logic control quite simply, which makes fuzzy logic a great tool to adapt it to different applications and to further develop them. A further advantage of fuzzy logic is the ease of how one can define a non-linear and complex control system in an intuitive way.

Principle of “fuzzy logic” based acquisition software

A fuzzy logic feedback control system was designed to control the automatic acquisition of MALDI spectra [2,3]. The implementation of these algorithms resulted in the AutoXecute™ software (Bruker) enabling unattended acquisition of MALDI spectra of peptides [4]. The control algorithm evaluates input parameters such as mass resolution and intensity of the base peak and then generates one or more output parameters, such as the laser fluence change or the decision to accumulate a series of spectra, based on a set of “rules” (see Fig. 2). Figure 3 shows the resulting parameter surface for peptide mapping experiments for the change of the laser attenuation as a function of intensity and resolving power of the base peak in a defined mass range.

After acquisition, spectra were automatically processed by macro scripts which are open to modifications by users. Standard routines include internal/external calibrations, peak annotations and printout, simultaneously with the acquisition

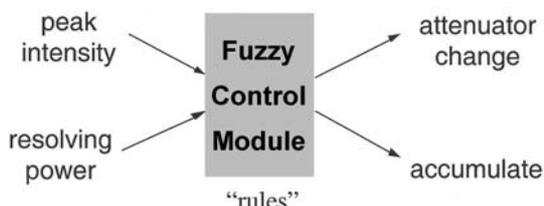


Figure 2. General set-up of a “fuzzy logic”-based control module. Output parameters (right) are calculated based on the input parameters (left) and a set of rules, like: “If peak intensity is low, change the attenuator for higher laser fluence”.

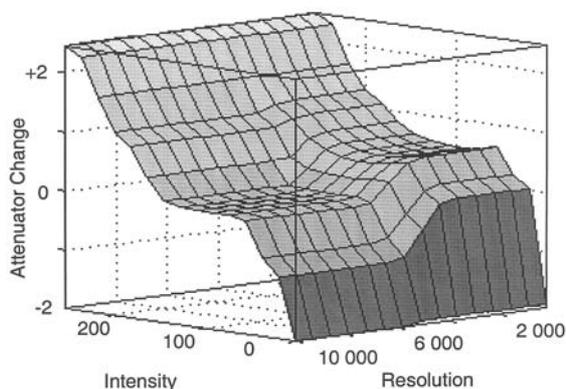


Figure 3. Parameter surface of the change of the laser attenuator as a function of peak intensity and resolving power (neg. attenuator changes increase the laser power).

of the next sample. Typical resolving powers were in the range between 5000 – 9000 at masses 1000 – 4000 Da and mass accuracies usually better than 40 ppm employing external calibration.

Beyond peptide applications, automated acquisitions of oligonucleotide spectra and proteins were carried out. The automatic acquisition of oligonucleotide spectra is of prime importance for genomic screening projects, where, e.g., PCR products are sized by MALDI either to detect gene mutations [5] or to analyse short tandem repeats (STR’s) as genetic markers.

Acquisition of MALDI-TOF spectra

All peptides and digests were prepared at 0.5 – 2 pmol in α -cyano-4-hydroxycinnamic acid matrix, nucleotides in 3-hydroxypicolinic acid. 0.5 μ L were prepared on the target and allowed to dry.

Spectra were all obtained in positive ion mode without operator interference on a REFLEX II instrument equipped with SCOUT™ ion source (26 sample positions) with video control, gridless 2-stage ion reflector, 1 GHz digitisation board and pulsed ion extraction accessory (Bruker Daltonik). Alternatively, a BIFLEX II with MultiProbe™ (10 sample positions) and video control was used (Bruker Daltonik). The software package consisted of AutoXecute™ 2.6 and XACQ™ 3.1 as data acquisition software and the XTOF™ 3.1 data processing package allowing processing to be defined in a script language (Bruker Daltonik). An acquisition parameter file was developed for good isotopic resolution in the mass range 1000 – 5500 Da, for all peptide mapping applications. Acquisition parameters were: 28.5 kV accelerating and 30 kV reflector voltage on the REFLEX and 19/20 kV on the BIFLEX.

The sample list set up for automatic measurements contains all information required to control the acquisition and the subsequent processing. This includes the number of shots to be accumulated or to be fired on a specific spot on the sample, target movement control, the application-specific selection of a fuzzy logic module, conditions when a sample is given up as “bad sample”, etc.

Processing of peptide spectra

Unless specified otherwise, spectra were processed using scripts, which start automatically subsequent to data storage.

The standard processing script “plot” displays the mass range 500 – 3500, labels all monoisotopic peaks beyond a certain threshold, writes a standard spectrum title and prints the annotated spectrum. More sophisticated scripts can internally calibrate on suitable peaks, remove known background signals, adducts and multiple charge states and perform a library search in intra- or internet search programs.

For instrument calibration, a peptide mixture in the mass range m/z 1050 – 3150 was prepared [4] and a calibration script was selected during the set-up of the sample list for automatic acquisition of this peptide mixture. It performs a

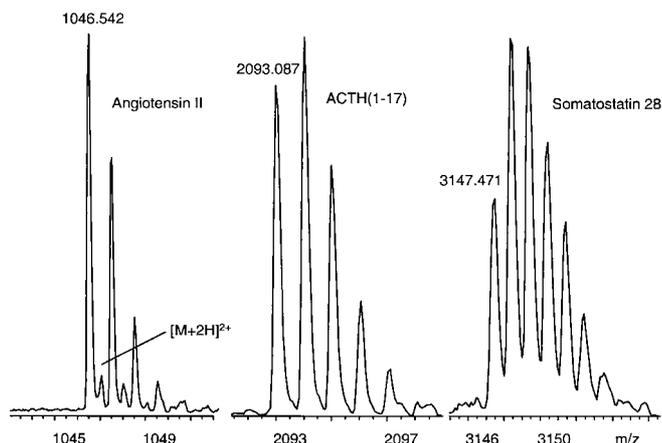


Figure 4. Details of the calibration peptide spectrum, which is used for completely automatic instrument calibration.

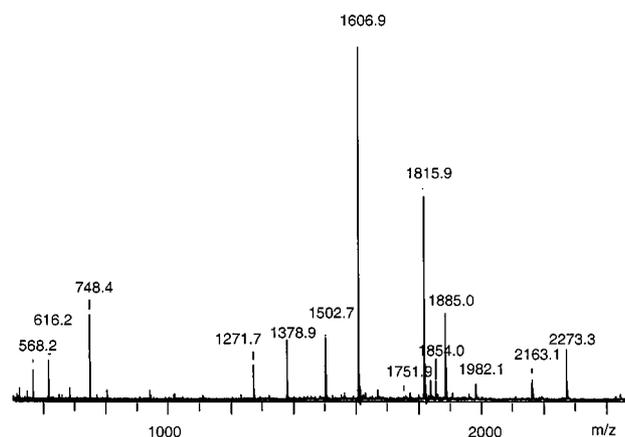


Figure 5. Tryptic peptide map of equine myoglobin; automatic acquisition, manual peak assignment. Data analysis in table I.

post-acquisition calibration of the data set which also calibrates the instrument for subsequent measurements. The expanded mass areas of three of these peptide calibrants in such spectrum is shown in figure 4. Interestingly, the $[M+2H]^{2+}$ peak of ACTH(1-17) is clearly resolved from the overlaying angiotensin II signal at m/z 1046, which illustrates the obtained resolution in automatically acquired spectra and that even doubly charged ions of peptides appear in PIE-MALDI spectra with significant intensity, which might interfere with automatic peak assignment routines if not considered.

The reproducibility of mass assignments across the 26 sample SCOUT target by automatic acquisition was also evaluated. On each position angiotensin II was prepared and spectra were automatically acquired. The mass errors were within ± 0.2 Da, which is typically the mass accuracy used for library searches if no internal calibration is used. For a further improvement of accuracy, each sample was internally calibrated by using known background signals like matrix or autoproteolytic fragments.

Library searches using protein mass fingerprints

Peptide mass information was obtained from a automatically acquired tryptic digest of horse myoglobin. All major peaks in the spectrum (Fig. 5) where manually assigned to horse myoglobin with a sequence coverage of 81% [4]. They gave an unchallenged high score for MYO_HORSE in a library search (Tab. I).

The peptide map of lysozyme was acquired and analysed completely automatic on the BIFLEX: annotation of all monoisotopic peaks was followed by scripts which eliminate known background signals (e.g., matrix, autoproteolytic fragments of trypsin) or adducts (e.g., sodiated peptide signals) and printout. Also the library search at the EMBL (http://www.mann.embl-heidelberg.de/Services/PeptideSearch/FR_PeptideSearchForm.html) or intranet searches were automatically performed after peak annotations. Peak

Table I. Experimental, theoretical MH^+ m/z values (from equine myoglobin) and absolute and relative mass errors for the main peaks in the spectrum in figure 4. External calibration with m/z 1046 and 2466 applies to masses with errors of determination of less than c.a. 40 ppm (shaded area).

<i>exp</i>	<i>calc</i>	<i>comment</i>	ΔDa	<i>ppm</i>
568.24	568.1356	3*CCA	0.10	183.75
616.24	616.18		0.06	97.37
659.39	659.38	Trypsin	0.01	15.17
684.37	684.37		0.00	0.00
748.43	748.44		-0.01	-13.36
1271.71	1271.66		0.05	39.32
1378.88	1378.84		0.04	29.01
1502.68	1502.67		0.01	6.65
1606.85	1606.86		-0.01	-6.22
1751.88	1751.92		-0.04	-22.83
1815.92	1815.903		0.02	9.36
1853.97	1853.96		0.01	5.39
1885.03	1885.02		0.01	5.30
1982.1	1982.06		0.04	20.18
2110.2	2110.15		0.05	23.69
2163.13	2163.06	Trypsin	0.07	32.36
2273.26	2273.16	Trypsin	0.10	43.99
		average	0.03	18

assignments of this spectrum are shown in figure 6 and the search results using ProFound in tables II and III. The sequence coverage in this case was 74% (without pre-sequence) and the other hits were strongly discriminated by their match probability.

Table II. Library search result from tryptic digest of reduced chicken lysozyme (from Fig. 5), using ProFound (<http://prowl.rockefeller.edu/cgi-bin/ProFound>).

Rank	Probability	Name	Long Name
1	1.0e+00	LYC_CHICK	LYSOZYME C PRECURSOR (CHICKEN)
2	4.6e-04	1LSG	lysozyme modified with fibrinogen gamma...
3	2.9e-04	1HEQ	lysozyme (EC 3.2.1.17) mutant (T40S, S91T)
4	1.1e-04	1HEN	lysozyme (EC 3.2.1.17) mutant (I55V, S91T)
5	3.1e-06	1LSN	lysozyme (EC 3.2.1.17) mutant (S91A)...
6	3.1e-06	2HFLY	lysozyme (EC 3.2.1.17) (with IgG1 Fab...
7	3.1e-06	1LSM	lysozyme (EC 3.2.1.17) mutant (I55L, S91T, ...
8	8.7e-13	GGLYS1_owl	GGLYS1 NID: g63580 - chicken.
9	1.3e-13	LYC_COLVI	LYSOZYME C (BOBWHITE QUAIL)
10	1.1e-13	LYC_LOPCA	LYSOZYME C (CALIFORNIA QUAIL)

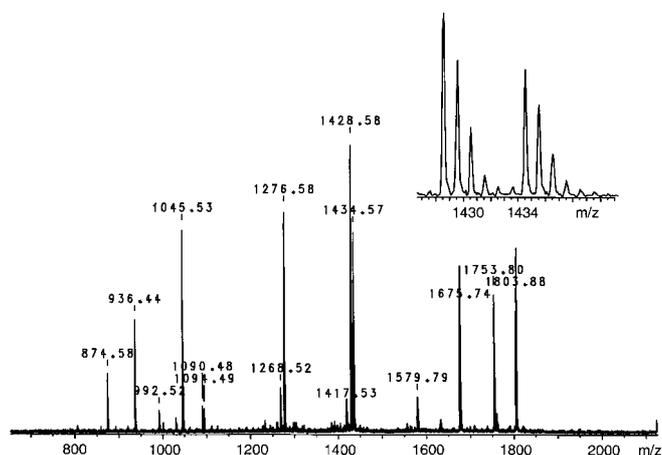


Figure 6. Tryptic peptide map of reduced chicken lysozyme; automatic acquisition, peak assignment and internet library search (see Tabs. II, III).

Table III. Identified peptide masses (neutrals) and sequence positions from the top scoring sequence in the search shown in table II.

Mass(exp)	Mass(calc)	Difference	From	To
873.57	873.41	0.16	15	21
935.43	935.37	0.06	62	68
991.51	991.50	0.01	6	14
1044.52	1044.54	-0.02	117	125
1267.51	1267.60	-0.09	22	33
1275.57	1275.64	-0.07	115	125
1427.57	1427.65	-0.08	34	45
1433.56	1433.63	-0.07	62	73
1674.73	1674.80	-0.07	98	112
1752.79	1752.83	-0.04	46	61
1802.87	1802.89	-0.02	97	112

Oligonucleotides

The acquisition of DNA spectra in the standard matrix 3-hydroxypicolinic acid is difficult compared to peptide spectra, because 1 μ L sample preparations are not homogeneous and "sweet spots" need to be found. Therefore, the AutoXecute™ software allows sample-specific movement controls and "spot hunting". However, this process is time consuming - it could require more than 2 min to obtain simple oligonucleotide spectra. A MALDI reflector mode spectrum of a 10 mer oligonucleotide (1 pmol) was obtained automatically (Fig. 7). In figure 8, a linear spectrum of an oligo-dT mixture (12 mer - 18 mer) is shown, where also the mass differences between peaks were assigned to the corresponding DNA base automatically. This indicates, that automatically acquired spectra can be used in the near future for applications which involve automatic sequencing of short reading length. Clearly, new sample preparation protocols

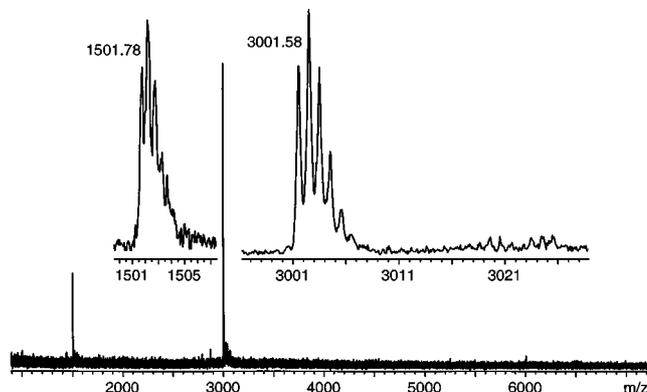


Figure 7. Reflector mode MALDI spectrum of DNA, 10 mer; automatic acquisition. Masses assigned to monoisotopic peaks of $[M+H]^+$ and $[M+2H]^+$. 1 pmol prepared in 3-hydroxypicolinic acid, 40 laser shots added.

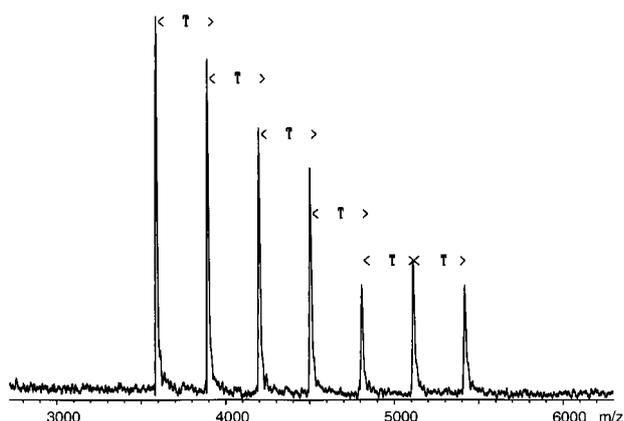


Figure 8. Linear mode MALDI spectrum of a mixture of oligo-dT. Acquisition and assignment of sequence labels was done automatically.

which involve a reduction of the spot size or provide homogeneous sample surfaces will greatly enhance the ability to automatically perform genetic analyses.

Conclusion

The developed fuzzy logic-based automatic acquisition of MALDI spectra is robust and enables the unattended acquisition of linear or reflector, continuous or pulsed ion extraction MALDI spectra of different analytes and masses.

Automatic acquisition of MALDI spectra is interesting for all applications which require the measurement of batches of similar samples, like mass determinations from chromatographic fractions (off-line LC-MS) or combinatorial che-

mistry. Mass accuracies and resolutions were comparable to spectra obtained by experienced operators at unrivaled speed (25 shot spectrum within c.a. 1 – 2 min including processing and library searches). This speed allows to obtain data from all 384 sample positions from new microtiter plate-format targets within c.a. 6 – 20 hours and currently allows a daily throughput of c.a. 400 – 1500 samples per day. Therefore, it is adequate to state "... that mass spectrometry provides the required throughput, the certainty of identification, and the general applicability to serve as the method of choice to connect genome and proteome" [2].

Although this throughput is much higher than even highly motivated persons would be able to achieve over several days, even higher throughputs will be required in large scale genetic screening projects. Besides improvements on sample preparation and specific fuzzy control algorithms for these applications, major future developments will include to create laboratory infrastructures which provide the capability to handle thousands of samples per day and to make automatic data analysis systems available, which require no user interference for data evaluation.

References

1. Brown, R. S.; Lennon, J. J. *Anal. Chem.* **1995**, *67*, 1998-2003.
2. Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. *PNAS* **1996**, *93*, 14440-14445.
3. Jensen, O.; Mortensen, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1997**, *69*, 1706-1714.
4. Suckau, D.; Kräuter, K.-O.; Rapp, U.; Mann, M.; Jensen, O. Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics, Palm Springs, CA, June 1-5, 1997; p 1057.
5. Little, D. P.; Braun, A.; Darnhofer-Demar, B.; Frilling, A.; Li, Y.; McIver, Jr. R. T.; Köster, H. *J. Mol. Med.* **1997**, *75*, 745-750.