Comparison of the possibilities of gas chromatography–mass spectrometry and tandem mass spectrometry systems for the analysis of anabolics in biological material

H.F. De Brabander\textsuperscript{a},* P. Batjoens\textsuperscript{a}, D. Courtheyn\textsuperscript{b}, J. Vercammen\textsuperscript{b}, K. De Wasch\textsuperscript{a}

\textsuperscript{a}Department of Veterinary Food Inspection, Laboratory of Chemical Analysis, Faculty of Veterinary Medicine, University of Ghent, Salisburylaan 133, B-9820 Merelbeke, Belgium
\textsuperscript{b}State Laboratory (ROL), Braemkasteelstraat 59, B-9050 Gentbrugge, Belgium

Abstract

Chromatographic techniques such as GC–MS play a most important role in modern multi-residue analysis of anabolic steroids. The major difference between GC–MS apparatus from different manufacturers is the way of detection and recording. Most apparatus use selected-ion monitoring (SIM) for the determination of low concentrations. Systems based on ion trap technology record in full-scan to even picogram concentrations using a computer algorithm to compare the most important peaks of the mass spectrum of the unknown to those of the standard.

In this investigation the possibilities of ion trap GC–MS and the recently released GCQ MS and MS\textsubscript{3} for the analysis of anabolics in biological material are compared.

Keywords: Mass spectrometry; Detection, GC; Anabolic agents; Steroids; Boldenone; Norethandrolone

1. Introduction

In Europe, the word “hormones” has a very bad reputation because of the possible danger for public health of some of these products, which are mostly anabolic steroids. Moreover, the consumer does not wish that animals for the meat market are treated with anabolic steroids, although toxicologists have declared that some are safe under certain application conditions. The Veterinary Food Inspection has to follow the legislation. However, a law that is difficult to apply is not of much value. Analytical laboratories have key functions in the control mechanisms because no legal action can be taken before the presence of residues of these products are proved with a high qualitative accuracy.

Chromatographic techniques play the most important role in modern multi-residue analysis of anabolic steroids, both for screening and for confirmatory analysis. In Belgium many HPTLC (high-performance thin-layer chromatography) methods are used, mainly in combination with a clean-up by solid phase extraction (SPE) (kidney fat) or HPLC (urine or faeces) [1–6]. However, most Belgian laboratories will confirm their HPTLC results systematically with GC–MS (gas chromatography–mass spectrometry) [7–9]. GC–MS apparatus from different manufacturers are used in Belgium. The major difference between them is the method of detection and recording of the chromatogram. Many apparatus use selected-ion monitoring (SIM) for the determination of low concentrations (<1 ng). Systems based on ion
trap technology as the Finnigan MAT Magnum and the Varian Saturn record in full-scan to even pg concentrations.

Recently, the latest and completely revised model of the Finnigan ion trap was released, i.e. the GCQ. This combination of quadrupole technology with the ultra-sensitive ion trap mass analyser gives the analyst new possibilities. The “tandem in time” system of this benchtop apparatus provides the added selectivity of MS–MS, where selected ions can be fragmented and the results analysed further. Our laboratories ordered the first GCQs in Europe and are building up experience in routine control. A first remark is that the nomenclature MS<sub>2</sub> is preferred to MS–MS because in the future further developments such as MS<sub>3</sub> (MS–MS–MS), MS<sub>4</sub>, etc. will be part of the possibilities for these types of machines.

In this investigation, the pros and cons of Magnum Ion Trap GC–MS and GCQ MS and MS<sub>2</sub> for the analysis of anabolics in biological material are compared using spiked samples. This comparison is focused on qualitative accuracy, i.e. the criteria that have to be fulfilled before the analyst may declare a sample positive. Possible quality criteria for GC–MS<sub>2</sub> are discussed. Quantitative analysis is of secondary importance for these illegal substances which have a so-called zero tolerance; it is impossible to quantify without a proper qualification. Moreover, in practice, the residue levels found are often below the limit of quantification. However, in the near future quantification in GC–MS<sub>n</sub> will become more and more important. In our laboratories, projects for proper quantification at the (sub) ppb level were started up.

2. Experimental

2.1. Apparatus

The following apparatus were used: Homogenisor (e.g. Waring Blendor with reservoir of 250 ml, Stomacher, Ultra-Turrax), microwave oven, centrifuge equipped with 300 ml centrifugation tubes, rotary vacuum evaporator, waterbath, extraction flasks (250 and 500 ml), vacuum manifold (e.g. Sample Preparation Unit, Analytichem, Harbor City, CA, USA), nitrogen evaporator (e.g. Techni Dry Block) or other types of evaporators (e.g. Speedvac SVC 200, SC 210A, Howe Gyrovap), chromatographic columns and tanks, autosampler vials [e.g. Chromacol 07-CPV (A)].

2.2. Reagents and reference components

Most reference steroids were obtained from Steraloids (St) (Wilton, NY, USA) or Sigma (Si) (St. Louis, MO, USA). Internal standards are equilenin (St E400) and spironolactone (St S200 or Si S3378). Other steroids were gifts from various sources. All recent standards were obtained through our NRL (National Reference Laboratory, IHE, Brussels) to ensure that all the field laboratories use the same standards [10]. The most important steroids used in this investigation are boldenone (BOL; 1,4-androstadien-17β-ol-3-one; St A200; M<sub>r</sub> 286.4), norgestrel (NG; 18,19-dinor-13β-ethyl-17β-hydroxy-4-pregnen-20-yn-3-one; Si N2260; M<sub>r</sub> 312.4), ethinylestradiol (EE2; 17α-ethynyl-1,3,5(10)-estratriene-3,17β-diol; Si E 4876; M<sub>r</sub> 296.4), fluoxymesterone (FMT; 9α-fluoro-11β-hydroxy-17α-methyltestosterone; Si F7751; M<sub>r</sub> 336.4) and norethandrolone (NE; 19-nor-4-androstien-17α-ethyl-17β-ol-3-one; St E3500; M<sub>r</sub> 302.4) MSTFA (N-methyl N-trimethylsilyltrifluoroacetamide) is from Macherey-Nagel (Düren, Germany) and TMSI (iodotrimethylsilane) is from Janssen Chimica (Geel, Belgium). dl-Dithiothreitol is purchased from Sigma-Chemie (Brussels, Belgium).

MSTFA<sup>+</sup> is prepared by dissolving 1% TMSI and a tip of a spatula point of the reductant dithiothreitol in MSTFA [9].

All solvents used were of analytical grade from Merck (Darmstadt, Germany).

2.3. Solutions

Stock solutions of anabolic steroids were prepared at 200 ng/μl in ethanol. Ten-fold dilutions of these stock solutions result in working solutions at a concentration of 20 ng/μl.

2.4. Methods

All clean-up methods used were described and validated before [9,11,12]. A short summary is as
follows: Urine (25 ml) is hydrolysed with Helix pomatia juice and extracted with ether. The crude extract is cleaned up with HPLC with fraction collection. Kidney fat (25 g) is extracted with methanol and after extraction of most of the fat with hexane, cleaned up with a two column SPE system. Faeces (20 g) is extracted with diethyl ether and after a two column SPE purification it is also cleaned up with HPLC with fraction collection.

2.5. Derivatisation

The final SPE extract or HPLC fractions are derivatised to TMS enol–TMS ether (trimethylsilyl) derivatives with MSTFA++. The sample or 10 μl of standard solution (200 ng) is transferred into an autosampler vial (700 μl) and dried under a nitrogen stream. MSTFA++ (25–50 μl) is added and the contents are mixed. A 1-μl volume (equivalent to 4 ng of standard) is injected into the GC.

2.6. GC–MS apparatus and conditions

A Magnum Ion Trap System (Finnigan MAT, San Jose, CA, USA) consisting of a Finnigan MAT A200S GC Autosampler, a Varian 3400 GC with 1077 capillary split/splitless injector, a Finnigan MAT Magnum Ion Trap Mass Spectrometer with electron impact and advanced positive chemical ionisation was used.

GC–MS conditions: Initial: 100°C to 250°C at 17°C/min, to 300°C at 2°C/min, ISO at 300°C for 3 min (total program ca. 37 min). Injector temperature, 260°C; transfer-line, 300°C.

Column: SGE BPX-5 (25 m×0.22 mm I.D., film thickness 0.25 μm).

Aquisition method: 1 scan/s over 25 min (mass range, 80–650 a.m.u.; filament-multiplier delay, 600 s). Tandem MS mode, 1 scan/s; mass range from 100 to a mass that was 1 unit higher than that of the parent ion selected (the parent ion is mostly the base peak of the full scan spectrum of the molecule); collision energy for fragmentation of the parent ion, 0.7–3.0 V.

3. Results and discussion

3.1. GC–MS of anabolics: present situation

In Belgium, anabolics are determined in different matrixes such as injection sites, excreta as urine and faeces and tissue samples such as kidney fat and meat. For all analytes, the clean-up procedure used for HPTLC is compatible with GC–MS (splitless injection). In most cases the remainder of the final product is derivatised with MSTFA++ (formation of TMS–enol ethers) [9].

In the laboratories using SIM, a number of diagnostic ions of the analyte are followed during a time window around the expected retention time. These diagnostic ions must be present in the correct relative intensities (±20% (CI) of ±10% (EI)). There is still a discussion about the number of diagnostic ions that must be followed. Two ions is certainly not enough; false positive results may be generated by isotope interference [13]. Four or more ions should be ideal from a theoretical point of view, but are not practicable at lower concentrations. From ringtests it was observed that the relative intensities of the ions do not remain constant enough when the concentration decreases [14]. Taking an objective analytical decision, three ions seems to be an acceptable compromise at this time.

The higher the number of ions the higher the specificity of the methods but also the higher the chance of false negative results when the identification criteria are strictly applied. The relative intensity of the ions may be disturbed by background noise and co-eluting substances and the sample has to be declared negative although the analyte is present.

Quality criteria for full-scan identification of low concentrations are not yet officially described. In most cases a computer algorithm compares the sixteen most important peaks of the mass spectrum of the unknown to those of the standard. In our
laboratory a FIT criterium of 800 (80% match) is used; the degree to which the library spectrum (standard) is included in the sample mass spectrum. For anabolics this is mostly fulfilled for higher concentrations (≥2 ppb). When the FIT is that high in full-scan apparatus, SIM conditions are mostly fulfilled in SIM machines. At lower concentrations, the analyte may disappear in a wood of interfering ions (Fig. 1) which lowers the FIT, although the SIM criteria may still be valid. This wood of interfering ions is caused by column bleeding and components from reagents and the matrix that are co-eluting with the analyte.

The situation presented in the theoretical example in Fig. 1 will happen with any analyte when its concentration is decreased. However, depending upon several circumstances, this effect may occur at different concentration levels.

A recent trend in GC–MS is the improvement of the detection limit in classical quadrupole apparatus. According to some manufacturers, lower concentrations may be measured in full scan so that both techniques may grow together in the future.

The present detection limits for anabolics in biological material with GC–MS may be calculated as follows: In the Finnigan Magnum the detection limit in full-scan of the derivate of an analytical standard is approximately 100 pg (0.1 ng). In the GC, ca. 1 μl of a 20-μl final extract can be injected. This corresponds to 2 ng extracted from 20 g of matrix. This corresponds with 0.1 ppb at a 100% yield of the clean-up. In practice, a detection limit of ca 0.5 ppb may be attained in full-scan.

In SIM much lower amounts of the same analytical standard may be detected (ca. 10 pg or 0.01 ng). With an injection of ca. 1 μl of a 20-μl final extract, this corresponds with 0.2 ng extracted from 20 g of matrix. The corresponding theoretical detection limit is 0.01 ppb (at a 100% yield of the clean-up). In SIM, GC–MS methods start from 2 g of material that brings the detection limit to ca. 0.1 ppb. However, in the determination of the real detection limit in GC–MS, factors other than the detection limit of standards are important, e.g., the mass range of the apparatus, the derivatisation and ionization method used (electron impact, positive or negative ionization) etc. It is not easy to compare apparatus from different manufacturers for different analytes in different matrices and in different modes.

In the preparation of the quality criteria for SIM and full-scan GC–MS mentioned above, most attention has been paid to obtaining as high a qualitative analytical accuracy as possible for a positive sample. In some cases (e.g. discussions, a second analysis, etc.) it is also very convenient to prove the absence of an analyte in a matrix. Next to the analysis of blank and spiked samples, the best way is the determination of the recovery of deuterated standards added to the sample at the concentration level of the detection limit (e.g. 2 ppb). Also, deuterated products are very useful in quantification. The big problem is the availability of these standards. The GC–MS data may also be combined with other results (e.g. HPTLC results). The identity of an analyte in a suspect sample may be based on two $R_f$ values in 2D-HPTLC, a characteristic colour or fluorescence after induction, a retention time window in HPLC used as clean-up, a retention time in GC$^2$ and a mass spectrum (full or diagnostic ions). These combinations of methods must fulfil the most stringent quality criteria.

3.2. Comparison of the Magnum and the GCQ

This comparison is based on the analysis of urine samples, spiked with some anabolics of current
interest at the level of the decision limit (2 ppb). The samples were cleaned-up and derivatised in routine procedures as described before [9,11,12]. They were injected in both the Magnum and the GCQ within a short time interval. The results for the quality parameters are summarized in Table 1.

As can be seen in Table 1, the FIT values (the results of the algorithm that measures how the sixteen most important peaks of the standard spectrum are included in the peaks of the sample spectrum) obtained in the Magnum are in most cases equal to, or greater than, 800 (our tentative criterium). In the GCQ, the FITS obtained are mostly higher than those found in the Magnum. This reflects a better match of the spectra with those of the spiked samples in the GCQ. The signal-to-noise ratio, measured from the sum of the three diagnostic ions of the anabolics is two to three-fold higher in the GCQ than in the Magnum.

The SIM parameters (the relative abundances of the three diagnostic peaks) measured on the background-subtracted spectra of the Magnum are not always in accordance with the 10% limit rule. In the GCQ, the difference between the ratios of the diagnostic peaks in sample and standard is less than that found in the Magnum, although some values still fall outside the theoretical limits (e.g. BOL, ion 430; 80 (for standard), −95 (for sample)).

Two anabolics are studied in more detail; BOL and NE. The mass spectra obtained at the correct retention time for the standard and the spiked urine sample for BOL and NE on the Magnum are shown In Fig. 2. The same spectra for the GCQ are presented in Fig. 3. The relative abundance of all the significant peaks of both anabolics on both apparatus are given in Table 2.

For the Magnum (Fig. 2), the spectra of both standards are easily recognized in the spectrum of the spiked sample, i.e., all the ions of the standards BOL and NE are present in the sample (see also Table 2). For BOL, the relative abundances of the ions 325, 415 and 430 are much lower in the sample than in the standard. This results in SIM criteria that are not fulfilled (although the analyte is present), but in FIT criteria that are intact. The FIT algorithm is more sensitive to the presence of peaks that match by mass rather than by intensity [15].

In the GCQ, the relative ratios of all the peaks are much more in concordance with each other, resulting in a nearly perfect fit between standards and spiked samples.

For norethandrolone, the biggest difference be-

<table>
<thead>
<tr>
<th>Analyte</th>
<th>FIT obtained (S/N)*</th>
<th>Ions</th>
<th>Magnum (relative abundance)</th>
<th>GCQ (relative abundance)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Standard</td>
<td>Sample</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Standard</td>
<td>Sample</td>
</tr>
<tr>
<td>BOL</td>
<td>954 (52)</td>
<td>206</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>966 (126)</td>
<td>430</td>
<td>63</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>325</td>
<td>44</td>
<td>16</td>
</tr>
<tr>
<td>NG</td>
<td>773 (32)</td>
<td>456</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>790 (48)</td>
<td>316</td>
<td>60</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>301</td>
<td>36</td>
<td>11</td>
</tr>
<tr>
<td>EE2</td>
<td>781 (99)</td>
<td>425</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>934 (184)</td>
<td>285</td>
<td>41</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>440</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>FMT</td>
<td>907 (12)</td>
<td>552</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>904 (78)</td>
<td>462</td>
<td>75</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>407</td>
<td>45</td>
<td>42</td>
</tr>
<tr>
<td>NE</td>
<td>915 (24)</td>
<td>446</td>
<td>45</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>971 (66)</td>
<td>287</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>356</td>
<td>27</td>
<td>29</td>
</tr>
</tbody>
</table>

*FIT = the results of the algorithm which measures how the sixteen most important peaks of the standard spectrum are included in the peaks of the sample spectrum; S/N = signal-to-noise ratio, measured as the sum of the three diagnostic ions of the anabolics.
Fig. 2. Mass spectra obtained for standards and urine samples spiked with boldenone and norethandrolone on the Magnum (2 ppb level).

The values presented above are measured manually on "background subtracted" spectra. They may differ from values obtained on crude spectra. Also, small variations may occur between successive background subtractions.

3.3. GC–MS–MS in residue analysis

GC–MS–MS, tandem MS or MS₂, have existed for some time on the "bigger" apparatus. However, in most cases these machines are too expensive for use in field laboratories performing residue analysis. Two years ago, a benchtop MS–MS based on a modification of an Ion Trap was introduced, i.e., the Varian Saturn. Instead of the classical MS–MS in space which consists of three mass spectrometers in series (one for the first mass analysis, one for a dissociation step and one for the second mass analysis), the whole MS–MS story takes place in one trap, as a function of time, controlled by a so-called MS–MS option. One ion (a precursor ion) is isolated in time and stored in the ion trap. Afterwards the dissociation of the precursor ion and the storage of product ions occurs in the same trap, but at a later time. The product ions are scanned from the trap at a third time, resulting in a product ion spectrum. In this way smaller and also cheaper apparatus may be constructed. However, this MS–MS system does not have all the capabilities of the bigger machines (e.g. parent ion and neutral loss scan). In contrast, ion traps theoretically allow MSⁿ, such that a mass spectrum from an ion of a mass spectrum from an ion of a mass spectrum etc. could be obtained in one run. Therefore, we prefer the nomenclature MS₂ instead of MS–MS.

In April 1994, Finnigan MAT also announced a MS–MS Magnum Performance Package upgrade for
the Magnum Ion Trap. However, this upgrade was not released on the market. Instead of an upgrade, a new and completely revised GC–MS–MS was presented, i.e. the GCQ. This apparatus is a combination of the well-known quadrupole technology with the ultra-sensitive ion trap mass analyser. Potentially, this apparatus could be a very powerful weapon in the battle against illegal abuse of anabolics in cattle fattening.

3.4. GC–MS$_2$ of anabolics on the GCQ

In our laboratory, GC–MS$_2$ of anabolics is not yet used as a technique on its own, but only in relation to GC–MS. The aim of GC–MS$_2$ spectra is to gain additional information to that already obtained in a previous GC–MS run. In Fig. 4 the MS$_2$ spectra of standards of BOL and NE and spiked urine samples taken at the correct retention time are presented.

For BOL, the MS$_2$ spectrum on parent ion 430 shows five important ions that are the same as those in the MS spectrum (with a slight difference between 191 and 189). However, less background, especially in the lower mass range is observed. For NE, the MS$_2$ spectrum on parent ion 287 shows eight
important ions that are different from those in the parent ion spectrum. Both MS$_2$ spectra taken on the sample match the standard MS$_2$ spectra taken up under the same conditions. This is also demonstrated in Table 3 where the intensities of all the peaks are given.

However, a MS$_2$ spectrum could be taken on various parent ions. In Fig. 5 MS$_2$ spectra for BOL and NE on the three diagnostic ions are shown.

It is obvious that in that way a lot of additional qualitative information about the analyte may be obtained. If all the MS$_2$ spectra of the sample match the MS$_2$ spectra of the standard, the risk of interference by co-eluting substances is almost negligible. On the GCQ, a chromatographic run for taking up each MS$_2$ spectrum on different parent ions is needed. In the future, MSn will be able to do the job in one single run with even more specificity; the successive MS$n$ spectra resulting from a previous daughter ion while in Fig. 5 all daughter spectra were obtained on the parent spectrum.

4. Conclusion

The evolution in the potential of analytical equipment, in their inter-relationship with the very fast developments in electronics and computers seems to be speeding up. This forces laboratories to continuously invest, both in equipment and in skilled personnel. For laboratory economics, it seems inevitable that the time used for clean-up has to be decreased and that the time spent on the final detection has to be increased. GC–MS$_2$ may play an important role in that philosophy. In this inves-
Fig. 5. MS$_2$ spectra of three of the diagnostic ions of boldenone and norethandrolone (standards).

tigation a part of the comparison between the Finnigan Magnum and the GCQ was presented. The GCQ is not only capable of successfully replacing the Magnum, but adds the identification power of MS$_2$. Next to that power, new features such as the extended mass range, NCI, ultra-SIM, etc. have to be explored. Anyway, it looks as if the use of a GCQ will lead us to shorter and more reliable analysis which may be repeated within an acceptable time limit.
Acknowledgments

The authors are indebted to Mrs. Ann Houvenaghel for skillful GC–MS operation of the Magnum.

References