Reprint from

RECENT ADVANCES
IN DOPING ANALYSIS
(8)

W. Schänzer
H. Geyer
A. Gotzmann
U. Mareck-Engelke
(Editors)

Sport und Buch Strauß, Köln, 2000

U. Flenker, V. Gougoulidis, W. Schänzer

Confirmation of Doping with Synthetic Endogenous Anabolic Steroids: Validation of the Method and Concepts of Inference

Institute of Biochemistry, German Sport · University Cologne

Abstract

Measurement of $^{13}C/^{12}C$-ratios of endogenous androgenic anabolic steroids allows for discrimination between biological or synthetic origin on principle. However a generally accepted method of data evaluation is missing up to now. A well suited method should be robust towards random variation of analytical techniques and towards diet dependent baseline differences while keeping discriminative power. In this paper we suggest a method based on classical methods of statistical model selection that in our opinion meets these requirements.

It is shown furthermore that apparent isotope ratios of endogenous steroids vary with sample preparation. Validity of the method is not restricted by this phenomenon, as reference and target compounds are affected in parallel.

1 Introduction

The measurement of ^{13}C-content of natural anabolic androgenic steroids (testosterone, androstendione, dihydrotestosterone, dehydroepiandrosterone, *etc.* ...) and/or their metabolites can help to elucidate the origin of these substances [1, 2, 4, 6, 10]. Synthetic material has been observed to be slightly depleted in ^{13}C, where the $\delta^{13}C$-values[1] of natural steroids roughly show a range from -18 ‰ to -25 ‰ and those of synthetic substances vary around -30 ‰.

In case of application of synthetic steroids the $\delta^{13}C$-values of the relevant analytes show a depletion of ^{13}C compared to some endogenous reference material not involved into the metabolic pathway under study. Hydroxylated pregnans are well suited reference substances in case of application of testosterone related steroids.

The measurement requires complete conversion of the steroids to CO_2. It is performed in a specially designed mass spectrometer [3, 8]. Briefly the ratio of $^{13}CO_2$ and $^{12}CO_2$ is determined following gas chromatographic seperation and quantitative combustion at 940 °C. For an extensive review of this method see [9].

However it still is unclear what difference of the $^{13}C/^{12}C$-ratio between target and reference compounds can be tolerated. Moreover one possibly has to take into account small

---

[1]$\delta^{13}C$ [‰] $= \left(\frac{(^{13}C/^{12}C)_{SPL}}{(^{13}C/^{12}C)_{STD}} - 1 \right) \times 10^3$

SPL: Sample; STD: Standard; International standard: PDB (Pee-Dee-Belemnite), $^{13}C/^{12}C = 0.0112372$.

differences of isotopic composition between different reference compounds on the one hand and probably considerable differences of baseline values between several samples on the other. Additionally the random variation of $^{13}C/^{12}C$-ratios caused by sample preparation and multiple measurements must be considered.

Here we suggest a procedure for data analysis that implicitly corrects for baseline effects both of biological and device dependent origin. The number of analytes may vary between several suspicous samples giving the method considerable flexibility. It is based on a widely accepted method of statistical model selection.

2 Theory

2.1 General Considerations

We basically make the following assumptions:

1. There is one sample suspected to contain synthetic anabolic androgenic steroids. This is the **test sample** SPL_T. There are steroids that either are applied themselves in their synthetic form or are metabolized to relevant substances in significant amount. These and their metabolites are the **target compounds**. Essentially no kinetic isotope effects occur during metabolism. Other classes of steroids do not contain synthetic material. These may serve as **reference compounds**. Target and reference compounds show sufficiently different isotopic composition.

2. A second sample exists, known to be free from synthetic steroids. This is the **reference sample** SPL_R.

3. Both samples are measured several times. At least one target compound and one reference compound are measured in SPL_T **and in** SPL_R. Generally an arbitrary number of compounds may be measured, **provided** these are identical with respect to the two samples.

Having performed all necessary measurements, we are in need to set up a model that explains our results. Basically two models can be formulated: A **null model** M_0 assuming no presence of synthetic steroids in the test sample and an **alternative model** M_1 taking into account isotopically different target compounds in SPL_T.

The decision of whether or not a positive case is present now depends on which model is justified to be selected.

Obviously M_0 is expected to cover the following observations:

1. All observed $\delta^{13}C$-values are equal within analytical precision both in SPL_T and SPL_R.

2. The $\delta^{13}C$-values of the target and the reference compounds are equal, but show differences in SPL_T and SPL_R (different "baseline"). This might be due to iso-
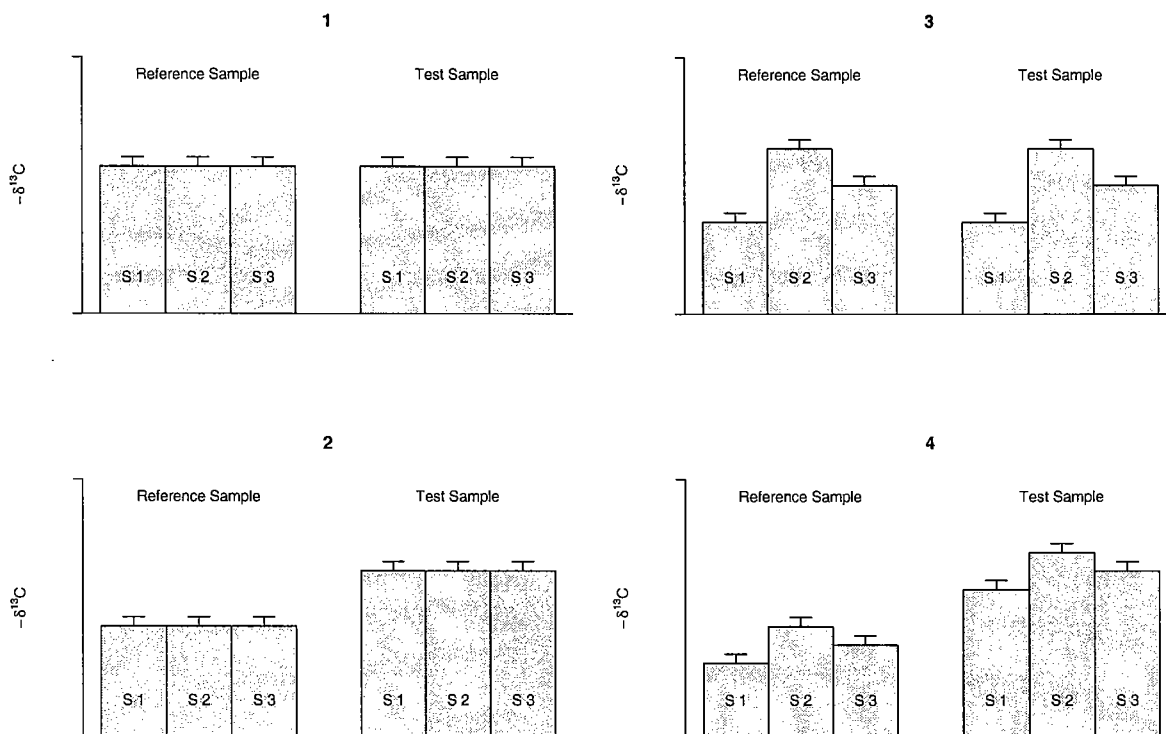
Figure 1: Result space compatible with M_0 (negative sample). S 1 – S 3 refer to different steroids and include reference and target compounds.

topically differing diets as well as to systematic error occuring during sample preparation.

3. There are differences between target and reference compounds or between different reference or target compounds or both. The **pattern** of these values is identical in SPL_T and SPL_R. This might be due to systematic analytical error – $i.e.$ certain substances wrongly show isotopic depletion or enrichment – or to isotope effects present in the metabolism of both SPL_T and SPL_R in similar magnitude.

4. Additionally SPL_T may show a common shift of all δ^{13}C-values $vs.$ SPL_R (or $vice$ $versa$) while the pattern itself is unaffected. It is likely then, that the two effects mentioned before show additive behaviour.

Schematically the result space compatible with M_0 is described by figure 1. For obvoius reasons M_0 can be refered to as a **parallel model**. If this model explains the data sufficiently, there is no reason to assume that any manipulationh as been performed. All values can be explained by effects that affect SPL_R in the same manner as SPL_T.

If we were able to show that the pattern of δ^{13}C-values in SPL_T differs from that of SPL_R **and** that our assumptions hold true, we are justified to conclude that additional effects are present. Most likely at least one of the analytes comes from a different source in SPL_T compared to SPL_R. Thus the inference to presence of exogenous steroids in

71

the test sample can be reduced to the question whether M_0 has to be rejected with given probability and M_1 has to be prefered.

2.2 Model Building

The most common strategy in statistical model building is to start from the model including as few effects as possible. Often one even will assume no effects (hence the term "null model"). The goodness of fit will be checked and additional effects will be added iteratively until no significant improvement of the fit is achieved anymore. If no significant effects are present in the data, M_0 will show sufficient fit and the algorithm stops at step one.

It is important to emphasize that the widely used t-statistic is just a different formulation for procedure described above [5]: The two-sample t-test is equivalent to the decision, whether one common or two different means should be assumed to describe the data inquired.

A generalized notation exists to formulate linear models [11], known as the "Wilkinson Rogers notation". Dependence of a variable from one ore more explanatory variables is denoted by a "$\sim$". If more than one explanatory variable is present, additive effects are symbolized by "+" and interactive effects are symbolized by "×". "Interaction" here means that the respective factor causes different effects when a second factor is present. Mathematically it is of no interest, whether the explanatory variables are continous – as in "usual" linear regression – or categorial. In the latter case "0" and "1" are introduced to code for presence or absence of a factor (For details see any textbook on linear reggression techniques, $e.g.$ [5, 11]).

Our two models described in 2.1 now can be written

$$M_0 \quad : \quad \delta^{13}\mathrm{C} \sim SPL + Steroid$$
$$M_1 \quad : \quad \delta^{13}\mathrm{C} \sim SPL \times Steroid$$

Here SPL has the two levels SPL_T and SPL_R and $Steroid$ refers to the different species of steroids that were measured.

The selection of the final model is managed via F-statistics: Any model will show a residual variance, $i.e.$ the variation not explained by the factors included in the model. Usually this can be interpreted as the analytical error. If this variation is reduced significantly when additional effects are introduced, the alternative model (M_1 in our case) has to be prefered. Otherwise M_0 is retained. Thus it is possible to determine the risk of wrongly rejecting M_0. This situation would be equivalent to a "false positive" case.

To show that the procedure is justified, it is essentially necessary to show, that analytical errors introduced by sample preparation and matrix effects behave purely additive with respect to the different steroids. In terms of model building the asumption of interaction between the factors "sample", "sample preparation" and "steroid" must not give rise to the goodness of fit.

3 Material and Methods

3.1 Samples

In order to establish the correctness of our basic assumptions from 2.1 the following urine samples were prepared and measured:

1. Ten routine urine samples testing negative for any kind of anabolic steroids.

2. One control urine collected from eight males known to be free from anabolic steroids (qualtity control urine, QC).

The routine samples were prepared three times. On each preparation three repeated measurements were performed.

For each preparation of a routine sample an aliquot of the QC was prepared in parallel. Each preparation of the QC underwent three repeated measurements as well.

In order to check the power of the established method, two samples suspicous for application of synthetic endogenous anabolic steroids were analyzed (T/E ratio > 6, high concentration of etiocholanolone and androsterone). For each of the suspicous samples another QC was prepared. Both QC's and boths suspicous samples were measured 5 times repeatedly.

3.2 Sample Preparation

The protocol sample preparation corresponds to a slightly modified "Screening IV" followed by HPLC cleanup. For details see [6].

3.3 GC/C/IRMS Measurements

Measurements were performed on a delta-C gas isotope ratio mass spectrometer (Finnigan MAT, Bremen, Germany) coupled *via* a "Combustion Interface II" (Finnigan MAT) to a 5890 gas chromatograph (Hewlett & Packard, Böblingen, Germany). Conditions of the whole device were kept as described in [6].

The analyzed compounds were 5β-pregnan-$3\alpha, 17\alpha$-diol (PD), 5β-pregnan-$3\alpha, 17\alpha, 20\alpha$-triol (PT), etiocholanolone and androsterone where a common value for the latter two compounds was calculated (EA).

Table 1: Model selection for the QC data

Model	Res. Df	Res. Sum Sq.	Df	Sum Sq.	F	p
M_0	87	92.70				
M_1	78	58.88	9	33.82	4.98	< 0.001
M_2	60	45.58	18	13.30	0.97	0.50

3.4 Model Building and Statistics

To asses effects of sample preparation and repeated measurements on δ^{13}C-values the following linear models were fitted using the QC data:

$$M_0 \quad : \quad \delta^{13}\text{C} \sim Steroid$$
$$M_1 \quad : \quad \delta^{13}\text{C} \sim Prep + Steroid$$
$$M_2 \quad : \quad \delta^{13}\text{C} \sim Prep \times Steroid$$

Steroid refers to the different steroid compounds analyzed (EA, PD, PT) and *Prep* refers to preparation numbers 1–10. The models were compared subsequently as described in 2.2.

As no evidence against additive effects of *Steroids* and *Prep* could be found (see 4), each QC urine was compared to its corresponding routine sample, were first an additive and then an interactive model was fitted. The data of the two suspicous samples were treated in the same manner. The formulas of the latter models are corresponding to those described in 2.2.

The software used was R-1.0.0 [7] running under Linux-2.2.10 on a personal computer[2].

4 Results and Discussion

The selection of an adequate model for the the data obtained from the QC urine is depicted in table 1. M_1 is clearly is best suited to describe the data. It shows significantly better fit than M_0 ($p < 0.001$). M_2 gives no significant improvement of the fit ($p = 0.50$). Therefore M_1 has to be retained.

A summary for this model is given in table 2. The following conclusions can be made:

1. Sample preparation causes significant alteration of apparent δ^{13}C-values in endogenous steroids (M_1 is more adequate than M_0).

2. The error introduced by sample preparation is equal for all compounds (M_2 does not fit better than M_1).

---

[2]R is free software. It is distributed under terms of the GNU license. It can be obtained from http://cran.r-project.org

Table 2: Summary of M_1

Coef.	Estimate	Std.err.	t	p
Prep. 1, EA	-24.35	0.32	-76.74	< 0.001
Prep. 2, EA	-23.97	0.32	-75.54	< 0.001
Prep. 3, EA	-24.39	0.32	-76.87	< 0.001
Prep. 4, EA	-24.57	0.32	-77.44	< 0.001
Prep. 5, EA	-22.40	0.32	-70.59	< 0.001
Prep. 6, EA	-23.35	0.32	-73.60	< 0.001
Prep. 7, EA	-23.30	0.32	-73.44	< 0.001
Prep. 8, EA	-23.46	0.32	-73.96	< 0.001
Prep. 9, EA	-23.72	0.32	-74.77	< 0.001
Prep. 10, EA	-23.81	0.32	-75.06	< 0.001
PD	-0.95	0.22	-4.22	< 0.001
PT	-0.18	0.22	-0.81	0.42

3. At least in the QC urine PD is depleted in ^{13}C vs. EA $(-0.95 \pm 0.22\ \%_{00}, p < 0.001)$.

4. PT shows similar $\delta^{13}C$-values as EA $(-0.18 \pm 0.22\ \%_{00}, p = 0.24)$

Obviously it is possible to maintain the concept of reference and test samples. Application of this method to the ten routine samples gave no positive result as can be seen in figure 2. In all cases it was possible to retain M_0 indicating that the three $\delta^{13}C$-values show more or less the same pattern in the QC and in the tested samples. There is no evidence that the different compounds are depleted or enriched in ^{13}C in different manner during sample preparation.

The frequent pattern of depleted ^{13}C content of PD compared to EA suggests that there is a slight isotope effect in the metabolism of this steroid. There was already evidence for this fact in the QC data. Nevertheless PD is suited as an endogenous reference compound as this is obviously a common phenomenon and as this can be covered by M_0.

The isotopic patterns for the two suspicous samples and the corresponding reference urines are represented graphically in figure 3. Whereas there are no objections against a parallel pattern in sample #30042, the sample #10109/99 clearly contains EA from an exogenous source. For sample #10109/99 M_0 has to be rejected with $p \sim 10^{-9}$. As is clear from the fact that sample preparation can cause systematic shift in $\delta^{13}C$-values of endogenous steroids, the absolute values are not suited very well as basis for inference. Inference based on reference and test samples should be fairly robust against this effect.

At the moment the reasons for the systematic parallel affection of $\delta^{13}C$-values by sample preparation are mostly speculative. Kinetic isotope effects of varying intensity depending on random effects could be an explanation.
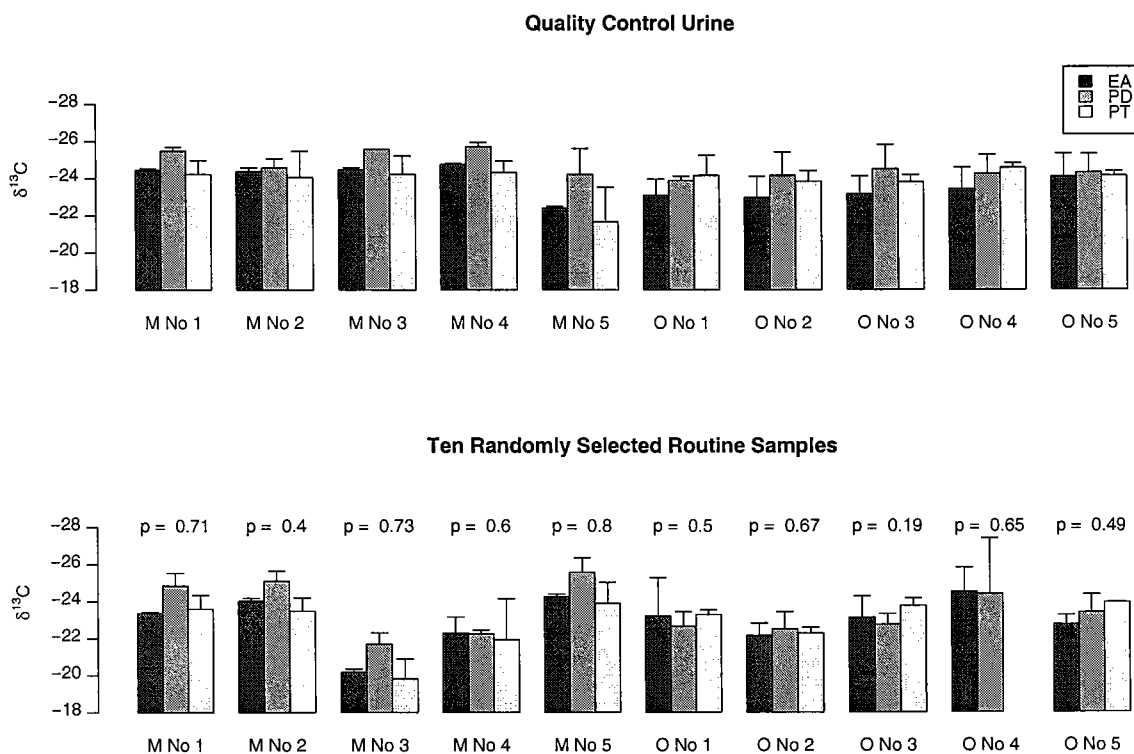
Quality Control Urine

Ten Randomly Selected Routine Samples

Figure 2: Means and standard deviations for δ^{13}C-values of EA, PD, and PT from ten routine samples and from one QC urine prepared ten times. Preparations were performed in parallel as indicated by labels. p-values refer to model selection according to 2.2

References

[1] R. Aguilera, M. Becchi, C. Grenot, H. Casabianca, and C. K. Hatton. Detection of testosterone misuse: comparison of two chromatographic sample preparation methods for gas chromatographic-combustion/isotope ratio mass spectrometric analysis. *J. Chromatography B*, 687:43 – 53, 1996.

[2] Roidrigue Aguilera, Michel Becchi, Hervé Casabianca, Caroline M. Hatton, Don H. Catlin, Borislav Starcevic, and Harrison G. Pope Jr. Improved method of detection of testosterone abuse by gas chromatography/combustion/isotope ratio mass spectrometry analysis of urinary steroids. *J. Mass Spec.*, 31:169–176, 1996.

[3] A. Barrie, J. Bricout, and J. Koziet. Gas chromatography-stable isotope ratio analysis at natural abundance levels. *Biomed. Mass Spec.*, 11(11):583–588, 1984.

[4] Michel Becchi, Rodrigue Aguilera, Yves Farizon, Marie-Magdeleine Flament, Hervé Casabianca, and Patrick James. Gas chromatography/combustion/isotope-ratio mass spectrometry analysis of urinary steroids to detect misuse of testosterone in sport. *Rapid Comm. Mass Spec.*, 8:304–308, 1994.

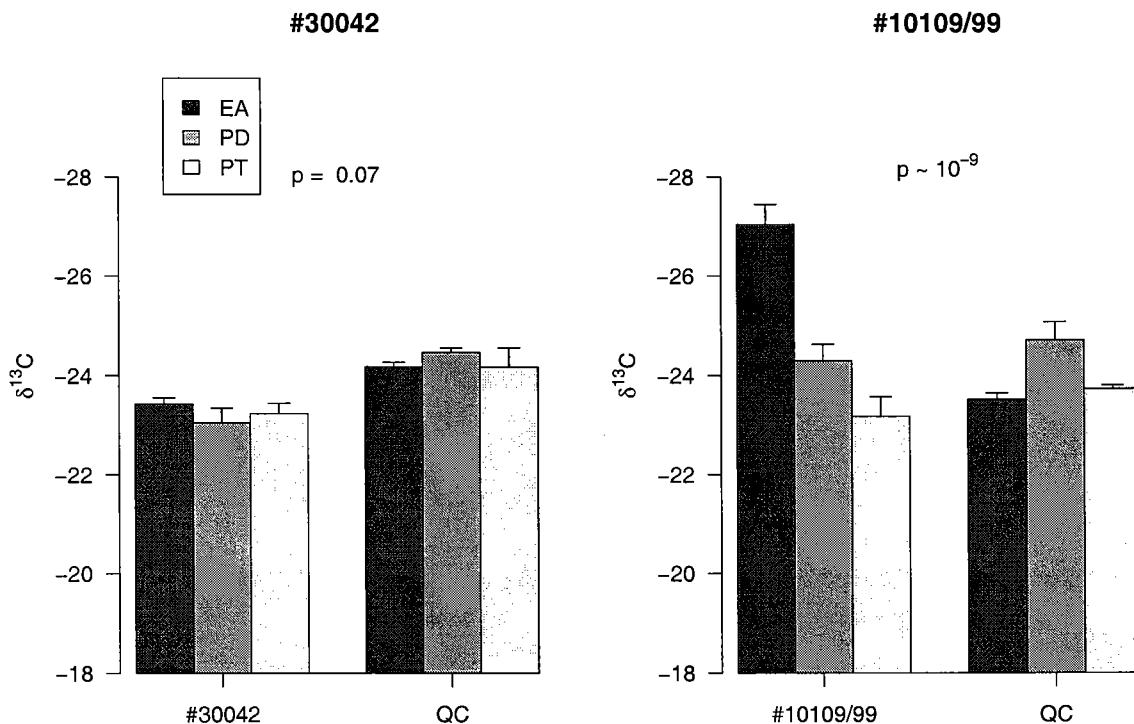[5] Annette J. Dobson. *An Introduction to Gneralized Linear Models*. Chapman & Hall, London, first edition, 1990.

Figure 3: Means and standard deviations for $\delta^{13}C$-values of EA, PD, and PT from two suspicous samples and corresponding reference samples. p-values refer to model selection according to 2.2

[6] Ulrich Flenker, Stevan Horning, Eckhart Nolteernsting, Hans Geyer, and Wilhelm Schänzer. Measurement of $^{13}C/^{12}C$-ratios to confirm misuse of endogenous steroids. In W. Schänzer, H. Geyer, A. Gotzmann, and U. Mareck-Engelke, editors, *Proceedings of the Manfred Donike Workshop*, number 6 in Recent advances in doping analysis, pages 243–256. Sport und Buch Strauß, Köln, 1999.

[7] Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314, 1996.

[8] D. E. Matthews and J. M. Hayes. Isotope-Ratio-Monitoring Gas Chromatography-Mass Spectrometry. *Anal. Chem.*, 50(11):1465 – 1473, September 1978.

[9] W. Meier-Augenstein. Applied gas chromatography coupled to isotope ratio mass spectrometry. *J. Chrom. A*, 842:351 – 371, 1999.

[10] Cedric H.L. Shackleton, Esther Roitman, Andy Phillips, and Tony Chang. Androstanediol and 5-androstenediol profiling for detecting exogenously administered dihydrotestosterone, epitestosterone, and dehydroepiandrosterone: Potential use in gas chromatography isotope ratio mass spectrometry. *Steroids*, 62(665–673), 1997.

[11] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S-PLUS*. Statistics and Computing. Springer, New York, Berlin, Heidelberg, third edition, 1999.