Journal of Chromatography A, 1216 (2009) 8404–8420

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Gas chromatographic quantitative structure–retention relationships of trimethylsilylated anabolic androgenic steroids by multiple linear regression and partial least squares

A.G. Fragkaki[a,b], A. Tsantili-Kakoulidou[c], Y.S. Angelis[a], M. Koupparis[b], C. Georgakopoulos[a,*]

[a] Doping Control Laboratory of Athens, Olympic Athletic Center of Athens "Spyros Louis", Kifisias 37, 15123 Maroussi, Greece
[b] Laboratory of Analytical Chemistry, Department of Chemistry, University of Athens, Panepistimioupolis, Zografou, 15771 Athens, Greece
[c] Department of Pharmaceutical Chemistry, School of Pharmacy, University of Athens, Panepistimioupolis, Zografou, 15771 Athens, Greece

ARTICLE INFO

Article history:
Received 25 May 2009
Received in revised form 8 September 2009
Accepted 25 September 2009
Available online 2 October 2009

Keywords:
Anabolic androgenic steroids
Doping control
Quantitative structure–retention relationships
Principal component analysis
Multiple linear regression
Partial least squares

ABSTRACT

A quantitative structure–retention relationship (QSRR) study has been performed to correlate relative retention times (RRTs) of trimethylsilylated (TMS) anabolic androgenic steroids (AAS) with their molecular characteristics, encoded by the respective descriptors, for the prediction of RRTs of novel molecules, using gas chromatography time-of-flight mass spectrometry (GC-TOF-MS). The elucidation of similarities and dissimilarities among the data structures was carried out using principal component analysis (PCA). Successful models were established using multiple linear regression (MLR) and partial least squares (PLS) techniques as a function of topological, three-dimensional (3D) and physicochemical descriptors. The models are useful for the estimation of RRTs of designer steroids for which no analytical data is available.

1. Introduction

Quantitative structure–retention relationships (QSRRs) represent a powerful technique for relating the gas chromatographic retention parameters of groups of analytes and their descriptors, which are quantities encoding the structural characteristics [1–3]. The most commonly used retention parameters in gas chromatography are the retention times (RTs), the relative retention times (RRTs), the Kováts retention indices and the logarithms of retention volumes of analytes [4]. The QSRR approach can be applied to identify the most useful structural descriptors, to predict retention for a new analyte, to gain insight into the molecular mechanism of chromatographic separation, to quantitatively compare separation properties of individual types of chromatographic columns and to evaluate properties other than chromatographic, such as lipophilicity. The construction of predictive QSRR models involves three steps [5]:

(a) the acquisition of a sufficiently large set of retention data of analytes covering possible structural diversities within a defined group of substances,
(b) the calculation of structural descriptors of the analytes, such as topological, three-dimensional (3D; geometrical and electronic) and physicochemical,
(c) the correlation of the retention data (dependent variable) with the calculated descriptors (independent variables) using appropriate statistical methods.

Multiple linear regression (MLR) is one of the most frequently applied methods in generating QSRR models [6]. The inability of MLR to treat intercorrelated variables and missing data, as well as the fact that it can consider only one dependent variable in each model can be overcome through partial least squares technique (PLS) which is also widely used in QSRR studies. Unlike MLR, PLS can analyze strongly collinear data, reducing the high dimensional data matrix to a much smaller and interpretable set of principal components or latent variables. Moreover, principal component analysis (PCA) is useful in providing a data overview [7].

Anabolic androgenic steroids (AAS) are included in the List of prohibited substances of the World Anti-Doping Agency (WADA) [8]. Chemically modified steroids, otherwise known as designer

* Corresponding author. Tel.: +30 210 6834567; fax: +30 210 6834021.
E-mail address: oaka@ath.forthnet.gr (C. Georgakopoulos).

0021-9673/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2009.09.066

steroids, that are circulated illegally in order to circumvent doping control, have been detected from doping control authorities. Currently there is a great effort from the WADA and WADA accredited laboratories to detect these illegal steroids in order to be included in the List of prohibited substances. AAS are extensively metabolized peripherally, notably in the liver and their target tissues, where conversion to their active forms is sometimes required before they can elicit their biological action [9,10]. However, excretion studies with illegal substances in humans are not allowed according to the Declaration of Helsinki.

Gas chromatography coupled to mass spectrometry (GC–MS) plays important role in doping control analyses identifying AAS and their metabolites as their trimethylsilyl- (TMS) derivatives using the electron ionization (EI) mode [11]. The evolution of time-of-flight mass spectrometry (TOF-MS) hardware, i.e., orthogonal acceleration and reflectron, has allowed its application to doping control analysis for accurate and sensitive full scan acquisition [12]. Consequently, the application of QSRRs for the prediction of RRTs of designer steroids and/or their metabolites using the GC-TOF-MS technique would be valuable. QSRR studies have been already performed in the doping control area for the prediction of the GC–MS RRTs of anabolic steroids [13], of stimulants and narcotics [14] and of α-, β_1- and β_2-agonists [15]. Other QSRR studies using steroid data have been also referred [16–18].

The aim of the present study was the development of QSRR statistical models of TMS-derivatized AAS that can be used for the prediction of GC-TOF-MS RRTs, following the results derived from our recent studies concerning the estimation of AAS metabolism schemes [9] and the AAS fragmentation patterns after TMS-derivatization [19], contributing significantly to the identification of designer steroids and their metabolites during doping control analyses.

Compared to the previous QSRR study related to the AAS RRT prediction [13], in the present study the number of compounds was expanded, while the comparison of the models derived by applying the MLR and PLS methods was performed. This QSRR study further estimated structural parameters that are involved in the chromatographic retention phenomena of the AAS molecules.

2. Experimental

2.1. Materials

The compounds included in the study were purchased from Sigma–Aldrich (Steinheim, Germany), Steraloids (Newport, USA) and LGC Standards (Wesel, Germany). Derivatization reagents were obtained as follows; N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) was purchased from Pierce (Rockford, USA), ammonium iodide (NH_4I) from Sigma–Aldrich (Steinheim, Germany), 1-propanethiol from Merck (Darmstadt, Germany) and acetonitrile (ACN) from Fischer Scientific Inc. (Leicestershire, UK).

2.2. Instrumentation

The chromatographic separation was performed on an Agilent (Santa Clara, USA) 6890N gas chromatograph and the TOF-MS analysis on a Waters Micromass GCT (Manchester, UK) orthogonal acceleration-reflectron TOF mass spectrometer in electron ionization mode at 70 Ev and at a mass resolution of 7000 FWHM. The GCT was operated with Waters Masslynx V4.SP4 software. The chromatographic column was Agilent Ultra 1, coated with methylsilicone gum, with a length of 17 m, internal diameter of 0.2 mm and film thickness of 0.11 μm. Helium was used as the carrier gas at a flow rate of 1.0 ml/min. The temperature program was: 70 °C initial temperature, set at 90 °C/min to 100 °C (held for 5 min) after which the oven is set at 30 °C/min to 180 °C and consequently set at 3 °C/min to 232 °C (held for 0 min). Finally the oven is set at 40 °C/min to 310 °C and the oven is held isothermally at 310 °C for 3 min. A 1 μl volume was injected in the split mode (1:15). The GCT averaged scans every 0.45 s with an interscan delay of 0.05 s and a mass range m/z 55–800. Data was acquired in the centroid mode.

Instrument tuning and calibration was performed using perfluorotributylamine (PCR Research Chemicals, Gainesville, FL, USA). Perfluorotributylamine was subsequently pumped out and replaced with 2,4,6-tris(trifluoromethyl)-1,3,5-triazine (also named metri, Aldrich, Milwaukee, WI, USA) for sample acquisition, in order to improve dynamic range of the GCT. During acquisition runs, metri was continuously introduced into the ion source.

2.3. Retention data generation

The GC-TOF-MS retention times (RTs) were determined for 64 TMS-derivatized AAS molecules and their metabolites. Their RRTs were identified on the basis of relative retention to compound methyl-testosterone di-TMS (internal standard) (Table 1). Chromatograms of representative analytes, included in the study, are shown in Fig. 1. It is also worth mentioning that no reversed elution order was observed for the analytes tested except for oxandrolone, possibly because it is structurally dissimilar with the majority of the rest analytes due to the existence of an heteroatom (oxygen) at ring A of its steroidal structure.

AAS molecules were TMS-derivatized following a two-step procedure:

- addition of 50 μl MSTFA and 25 μl ACN at 80 °C for 30 min and then,
- addition of 50 μl MSTFA/NH_4I/propanethiol 1000:4:7 v:w:w at 80 °C for 30 min.

2.4. Descriptor generation

Diverse topological, 3D and physicochemical descriptors were calculated for the present QSRR study (81 descriptors for each analyte of the dataset). Topological descriptors included molecular connectivity descriptors, shape indices κ (kappa) and path descriptors [20]. Molecular connectivity descriptors and κ indices were calculated from the hydrogen-suppressed graph of the molecule, encoding information about the degree of branching and size of the molecules. The hydrogen-suppressed graph is a theoretical representation where each structure is shown as a graph consisting of vertexes (atoms) and edges (bonds). The calculation of electrotopological state indices also assumed the hydrogen-suppressed graph to represent the structure of the molecule, combining both the topological environment and the electronic character of each skeletal atom in a molecule. The values of topological and electrotopological descriptors were calculated using the MOLCONN-Z v.4.0 program (Quincy, USA) [21].

The 3D descriptors were calculated after the minimization of the strain energy of the molecular structure and the calculation of the three-dimensional coordinates [4]. The energy minimization was achieved using the MM2 algorithm and the semiempirical molecular orbital MOPAC algorithm with the Austin model 1 (AM1) Hamiltonian of Chem3D option included in the ChemOffice Ultra 6.0 molecular modeling package (CambridgeSoft, Whitehouse, USA) [22]. 3D descriptors include the solvent-accessible surface areas and several types of the molecular energy, such as stretch, torsion, bend, stretch-bend, 1,4-van der Waals, dipole–dipole energy, etc. Furthermore, 3D descriptors include the electronic ones, which encode information about the electronic environment of each molecule and are ranked into two categories, global and local.