Journal of Steroid Biochemistry & Molecular Biology 126 (2011) 35–45

Contents lists available at ScienceDirect

Journal of Steroid Biochemistry and Molecular Biology

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

Anabolic and androgenic activities of 19-nor-testosterone steroids: QSAR study using quantum and physicochemical molecular descriptors

Yoanna María Álvarez-Ginarte[a], Luis Alberto Montero-Cabrera[a], José Manuel García de la Vega[b,*], Pedro Noheda-Marín[c], Yovani Marrero-Ponce[d], José Alberto Ruíz-García[e]

[a] Laboratory of Theoretical and Computational Chemistry, Faculty of Chemistry, University of Havana, 10400 La Habana, Cuba
[b] Departamento de Química Física Aplicada, Universidad Autónoma de Madrid, 28049 Madrid, Spain
[c] Institute of Organic Chemistry (IQOG), Superior Council of Scientific Investigations (CSIC), 28006 Madrid, Spain
[d] Unit of Computer-Aided Molecular "Biosilico" Discovery and Bioinformatic Research (CAMD-BIR Unit), Faculty of Chemistry-Pharmacy, Central University of Las Villas, Santa Clara 54830, Villa Clara, Cuba
[e] Biomolecular Chemistry Center, 16042 La Habana, Cuba

ARTICLE INFO

Article history:
Received 13 October 2010
Received in revised form 29 March 2011
Accepted 3 April 2011

Keywords:
Anabolic activity
Androgenic activity
Quantum molecular descriptor
Physicochemical molecular descriptor
Quantitative structure–activity relationship
Testosterone family and cluster analysis

ABSTRACT

Quantitative structure–activity relationship (QSAR) study of 19-nor-testosterone steroids family was performed using quantum and physicochemical molecular descriptors. The quantum–chemical descriptors were calculated using semiempirical calculations. The descriptor values were statistically correlated using multi-linear regression analysis. The QSAR study indicated that the electronic properties of these derivatives have significant relationship with observed biological activities. The found QSAR equations explain that the energy difference between the *LUMO* and *HOMO*, the total dipole moment, the chemical potential and the value of the net charge of different carbon atoms in the steroid nucleus showed key interaction of these steroids with their anabolic–androgenic receptor binding site. The calculated values predict that the 17α-cyclopropyl-17β, 3β-hydroxy-4-estrene compound presents the highest anabolic–androgenic ratio (AAR) and the 7α-methyl-17β-acetoxy-estr-4-en-3-one compound the lowest AAR. This study might be helpful in the future successful identification of "real" or "virtual" anabolic–androgenic steroids.

1. Introduction

Anabolic/androgenic steroids (AAS) exert two different effects: development and maintenance of secondary male sexual characteristics (androgenic effects) and promotion of muscle growth (anabolic effects) [1]. These drugs are used in the fast recovery from protein-wasting disorders. In HIV patients, anabolic steroids are used to regain lean muscle mass, as well as to prevent organ failure and secondary immune dysfunction [2]. These compounds have proved to be an effective oral therapy to promote weight gain after extensive surgery, chronic infections and severe trauma [3]. They are indicated in the treatment of anemia caused by deficient red-cell production [4], osteoporosis [5] and metastatic cancer [6].

Generally, testosterone is considered to be a poor AAS because, when orally taken, it is rapidly degraded, and only small amounts reach the systemic circulation. Additionally, when testosterone is injected, effective levels of the drug are not sustained because of rapid degradation [7]. In order to maximize the effectiveness of AAS, the basic chemical structure of testosterone is generally modified by (I) esterification at the 17β-hydroxyl group, (II) alkylation at the 17 α-position, or (III) modification in any of the 1, 2, 9, or 11 carbon of the ring structure in the molecule (see Fig. 1) [7].

Family of 19-nor-testosterone is produced by the removal of the 19 methyl group of testosterone (see Fig. 1). This change increases the anabolic effect. The 19-nor-testosterone is more planar than testosterone, increasing its receptor-binding affinity. However, it binds with similar affinity to androgen receptors in skeletal muscle and in prostate. The 19-nor-testosterone is not converted to dihydrotestosterone (DHT) in tissues containing 5α-reductase enzyme. Instead, it is converted to dihydro-19-nor-testosterone that is a compound that binds less tightly to androgen receptor in the same way that the DHT. These factors could explain the diminished effect of 19-nor-testosterone relative to testosterone on androgen target tissues containing 5α-reductase enzyme (e.g., seminal vesicles) together with a greater effect than testosterone on tissues containing little or none of this enzyme, e.g., levator ani and skeletal muscle [8].

In an attempt to meet these goals many AAS drugs have been created with various levels of success. However, there is no AAS developed to date that contains a pronounced anabolic potency, while having no androgenic effect [9]. It is important to note that with the creation of new AAS, subsequent modifications result in

* Corresponding author. Tel.: +34 91 497 4963; fax: +34 91 497 4512.
E-mail address: garcia.delavega@uam.es (J.M.G. de la Vega).

0960-0760/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsbmb.2011.04.003

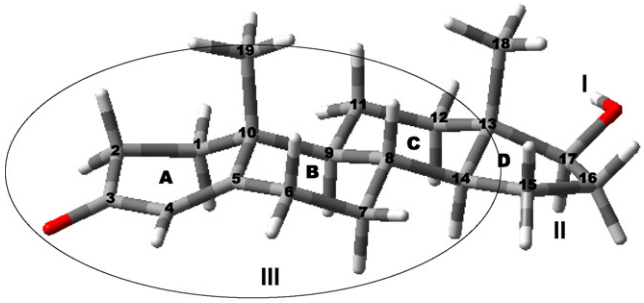


Fig. 1. Three types of modifications of testosterone used to produce AAS. Modification I: esterification at the 17β-hydroxyl group. Modification II: alkylation at the 17 α-position. Modification III: modification in any of the 1, 2, 9, or 11 carbon atoms in the steroid nucleus.

alterations in the drug's anabolic to androgenic ratio and modify the potential side effects that could occur in response to the drug. In integrating both measures, the anabolic index, which relates the ratio of anabolic to androgenic response for a given steroid, is used. If the anabolic index is higher than one, it indicates a higher trend towards anabolic effect and, therefore, classifies the drug as an anabolic steroid. A measure lower than one, in turn, assesses the steroid as androgenic [10].

Several authors have developed QSAR models for estrogens [11], progestagens [12], and corticosteroid [13] steroid hormones. The steroid benchmark with the corresponding globulin affinity is the most extensively studied dataset of steroids [14–18]. Recently, we reported multi-linear regression (MLR) QSAR models for congeneric series of AAS: 17β-hydroxy-5α-androstane [19], 4,5α-dihydrotestosterone [20] and testosterone [21] steroid families. We reported too a robust *biosilico* model of linear discriminant analysis (LDA) [22]. This model was used to analyze the anabolic–androgenic activities of structurally diverse steroids and to discover novel AAS, as well as to give a structural interpretation of their AAR. We have selected 366 steroids with a structural variability and included four different steroids families: 17β-hydroxy-5α-androstane, 4,5α-dyhidrotestosterone, testosterone and 19-nor-testosterone derivatives [23]. The general idea of our work is to develop general models of classification for steroids with high and moderate-low anabolic–androgenic ratio (AAR). Subsequently, we quantify their anabolic and/or androgenic activities in the models of MLR according to the family that each selected molecule belong to. Finally, our approach could help in the future successful identification of "real" or "virtual" AAS.

The main aim of the present report is to develop MLR QSAR models for anabolic and androgenic steroids of the 19-nor-testosterone steroids family using quantum and physicochemical MDs and a genetic algorithm (GA) as a method for the selection of the best set of variables. The QSAR models provide a structural interpretation of their anabolic–androgenic activities and design new anabolic and androgenic steroids.

2. Methods

2.1. The studied compounds and their biological activity data

In the present work, logarithm of the inverse of biological activity was used in order to establish QSAR equations. It was taken from literature a data set of 26 compounds derived from 19-nor-testosterone steroids family with anabolic and androgenic activities determined *in vivo*. In pharmacological activity determination were used thirty male Wistar rats (21 ± 23 days old, body weight: 50–90 g) maintained seven days with food and water available *ad libitum*. In the determination of the anabolic and androgenic activities, researchers isolated three organs from each rat: the seminal vesicle (SV), the ventral prostate (VP), and the levator ani muscle (LA) [22].

These organs were all weighted and a comparison between the active groups and the placebo groups was made. The differences in weight of the seminal vesicles and the ventral prostates represent the androgenic activity, while the difference in weight of the levator ani muscle between the control and active group represents the anabolic activity [22]. The experimental values of these biological activities and molecular structures for all steroids are shown in Table 1 and Fig. 2, respectively.

The data set was divided randomly into three training sets ($n = 21$, 80% of the data): (1) steroids with anabolic activity, expressed by log (1/LA); (2) steroids with androgenic activity, expressed by log (1/VP) and by log (1/SV) and (3) AAR values, expressed by [log (1/LA)/log(1/VP)] and [log (1/LA)/log(1/SV)]. The predictive ability of each model was then evaluated by test sets including the remaining steroids ($n = 5$, 20% of data): (1) test set for anabolic activity; (2) test set for androgenic activity and (3) test set for AAR. The experimental values of AAR were quantified using the anabolic and both androgenic activities values shown in Table 1.

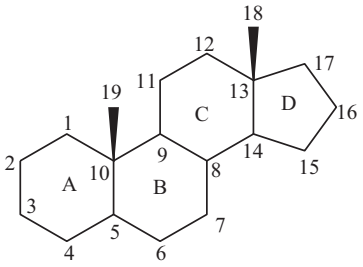
2.2. Calculated 2D-descriptors

A large number of MDs are usually used in QSAR methods [24]. The specific biological action of drugs is frequently described by hydrophobic, electronic and steric properties. The hydrophobic properties express the ability of a molecule to be transported through the organism in order to interact with biological membranes and to be bound to the receptor by van der Waals forces. We considered as hydrophobic descriptor the logarithm of the octanol–water partition coefficient (log *P*) [25]. Electronic and steric properties characterize the pharmacodynamic properties in the ligand–receptor interaction. They define the ability of the drug to join the receptor [26]. Calculated electronic descriptors by quantum mechanical procedures were: hydration energy (E_{H_2O}) [27], polarizability (*P*) [28], the total dipole moment (μ) of the molecule, electronic energy (*E*), total energy (E_T), *HOMO* (highest occupied molecular orbital) eigenvalue, *LUMO* (lowest unoccupied molecular orbital) eigenvalue, energy difference between the *LUMO* and *HOMO* (ΔE_{L-H}), net atomic charges of C atoms 1–17 in the steroid backbone (q_1–q_{17}), chemical hardness (η), softness (*S*), chemical potential (*U*) and electrophilicity index (ω) [29,30]. Electronic descriptors were calculated with the parametric method 3 (PM3) semi-empirical Hamiltonian [31] after the full geometrical optimization of each molecule using MOPAC 6 software [32]. The steric properties analyses were: approximate surface area (ASA), grid surface area (GSA) [33], molar volume (MV) and molar refractivity (MR) [34]. The MDs calculated in the present work and that were included in QSAR models are given in Table 2. Correlations among physicochemical parameters are listed in Table 3.

2.3. Statistical analysis

The BuildQSAR software [35] was employed to perform variable selection and QSAR modeling. The mutation probability was specified as 35%. The length of the equations was set for three or four terms (according to the models sought-after) and a constant. The population size was set to 100. The GA with an initial population size of 100 rapidly converged (200 generations) and reached an optimal QSAR model in a reasonable number of GA generations [36–38]. The search for the best model can be processed in terms of the highest determination coefficient (R^2), *F*-test value, *p*-value ($p < 0.01$) and the lowest standard deviation (s) [39]. The validity of model was examines using the leave-one-out cross validated procedure. From this procedure, we define the predictive squared

Table 1
Numbering of the steroidal backbone as well as the experimental values of anabolic and androgenic activities for 19-nor-testosterone derivatives.



Compound[a]	log (1/LA)[b]	log (1/VP)[c]	log (1/SV)[c]
1. 19-Nor-testosterone	2.10	1.70	1.70
2. 19-Nor-testosterone-propionate	1.94	1.79	1.79
3. 17α-Methyl-19-nortestosterone	1.95	1.93	1.93
4. 17α-Ethynyl-19-nortestosterone	1.60	0.48	0.48
5. 17α-Ethyl-17β-hydroxyestr-4-ene	2.30	1.60	1.60
6. 7α-Methyl-17β-acetoxy-estr-4-en-3-one	2.22	2.22	2.22
7. 7α,17α-Dimethyl-17β-hydroxyestr-4-en-3-one	2.51	2.34	2.34
8. 4-Chloro-17α-methyl-17β-hydroxyestr-4-en-3-one	1.85	1.40	1.40
9. 4-Chloro-17α-methyl-17β-propionoxyestr-4-en-3-one	1.85	1.15	1.15
10. 4-Chloro-17β-acetoxyestr-4-en-3-one	1.89	0.48	0.48
11. Estr-4-en-3,17-dione	1.76	1.65	1.65
12. 7α-Methylestr-4-en-3,17-dione	2.09	2.08	2.08
13. 16β-Methyl-17β-hydroxy-estr-4-en-3-one	1.60	1.00	1.00
14. 2α-Methyl-17β-hydroxy-5α-estran-3-one	1.15	0.85	0.85
15. 17α-methyl-3β,17β-dihydroxyestr-5-ene	2.85	1.70	1.70
16. 17α-Ethyl-3β,17β-dihydroxyestr-5-ene	2.60	1.30	1.30
17. 17β-Hydroxy-5α-estr-2-ene	1.30	0.30	0.30
18. 17β-Acetoxy-Δ^4-estreno[2,3-d] isoxazole	1.90	1.00	1.00
19. 6α-Methyl-17β-hydroxy-estr-4-en-3-one	1.41	1.23	1.23
20. 17β-Hydroxy-estra-4,9(10)-dien-3-one	2.00	1.00	1.00
21. 17α-Methyl-17β-hydroxy-estra-4,9(10)-dien-3-one	2.11	1.48	1.48
22. 17α-Ethyl-17β-hydroxyestra-4,9(10)-dien-3-one	2.00	1.00	1.00
23. 17β-Acetoxyestra-4,9,11-trien-3-one	2.70	2.70	2.70
24. 17β-Methoxy-methyloxyestra-4,9,11-trien-3-one	3.30	3.30	3.30
25. 17β-Hydroxyestra-4,9,11-trien-3-on-17-n-undecanoate	2.70	2.70	2.70
26. 17α-Cyclopropyl-17β,3β-hydroxy-4-estrene	1.70	0.78	0.78

[a] Structure of compound give in Fig. 2.
[b] Anabolic activity expressed by the inverse logarithm of the levator ani muscle (LA) weight.
[c] Androgenic activity expressed by the inverse logarithm of the ventral prostate (VP) and seminal vesicle (SV) weights, respectively.

Table 2
Quantum and physicochemical parameters values included in the QSAR models for steroids in database.

Compound[a]	ΔE_{L-H}[b]	q_6	q_7	q_9	q_{11}	q_{12}	q_{13}	U[c]
1	9.99	−0.07	−0.09	−0.07	−0.11	−0.08	−0.05	−5.09
2	10.00	−0.07	−0.09	−0.07	−0.11	−0.08	−0.05	−5.16
3	9.99	−0.07	−0.09	−0.07	−0.11	−0.09	−0.04	−5.08
4	9.99	−0.07	−0.09	−0.07	−0.11	−0.09	−0.04	−5.08
5	10.46	−0.05	−0.09	−0.07	−0.11	−0.09	−0.04	−4.11
6	9.98	−0.07	−0.07	−0.08	−0.11	−0.08	−0.05	−5.14
7	9.97	−0.07	−0.07	−0.08	−0.11	−0.09	−0.04	−5.06
8	9.13	−0.08	−0.09	−0.07	−0.11	−0.09	−0.04	−4.88
9	9.13	−0.08	−0.09	−0.07	−0.11	−0.09	−0.04	−4.88
10	9.13	−0.08	−0.09	−0.07	−0.11	−0.08	−0.05	−4.95
11	10.00	−0.07	−0.09	−0.07	−0.11	−0.07	−0.10	−5.19
12	9.98	−0.07	−0.07	−0.08	−0.11	−0.07	−0.10	−5.17
13	10.00	−0.07	−0.09	−0.07	−0.11	−0.08	−0.08	−5.12
14	11.26	−0.10	−0.10	−0.07	−0.11	−0.08	−0.05	−4.75
15	10.46	−0.15	−0.06	−0.08	−0.11	−0.09	−0.04	−4.26
16	10.46	−0.15	−0.06	−0.08	−0.10	−0.09	−0.04	−4.25
17	10.75	-0.10	−0.10	−0.07	−0.11	−0.08	−0.05	−4.22
18	8.44	−0.06	−0.10	−0.07	−0.11	−0.09	−0.05	−4.87
19	10.00	−0.04	−0.10	−0.07	−0.11	−0.08	−0.05	−5.07
20	8.57	−0.07	−0.10	−0.07	−0.07	−0.08	−0.05	−4.80
21	8.56	−0.07	−0.10	−0.07	−0.07	−0.09	−0.04	−4.79
22	8.56	−0.07	−0.10	−0.07	−0.07	−0.09	−0.04	−4.79
23	8.06	−0.07	−0.10	−0.04	−0.12	−0.11	−0.03	−4.93
24	8.06	−0.07	−0.10	−0.04	−0.11	−0.12	−0.06	−4.89
25	8.05	−0.07	−0.10	−0.04	−0.12	−0.10	−0.03	−4.87
26	10.51	−0.06	−0.09	−0.07	−0.11	−0.09	−0.04	−4.36

[a] Number of these compounds and structures are given in Table 1 and Fig. 2, respectively.
[b] Energy difference between the *LUMO* and the *HOMO*. *q*, net atomic charges of carbons 6, 7, 9, 11, 12 and 13, respectively in the steroids backbone.
[c] Chemical potential.

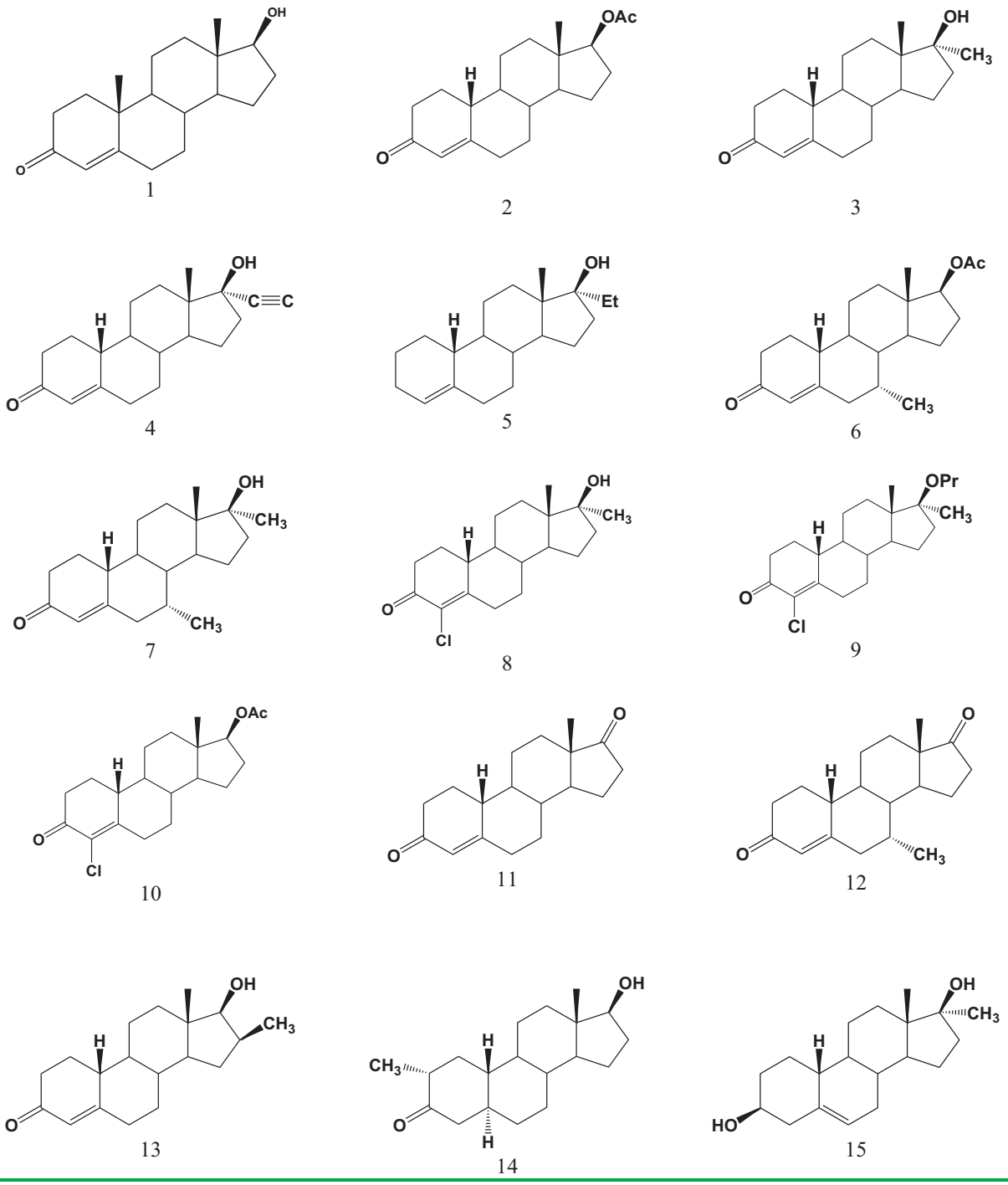


Fig. 2. The structures of the 19-nor-testosterone family considered in the data set of this study.

correlation coefficient (q^2) according to the following expression:

$$q^2 = 1 - \frac{\sum(yi - \widehat{yi})^2}{\sum(yi - \bar{y})^2} \quad (1)$$

where yi are the observed activities, $\widehat{yi}$ are the estimates activities by the models and $\bar{y}$ is the average activity [40,41]. Many authors consider high q^2 values (for instance, $q^2 > 0.5$) as an indicator or even as the ultimate proof of the high predictive power of a particular QSAR model [40,41]. Nevertheless, Golbraikh and Tropsha have recently demonstrated that high values of q^2 appear to be a necessary, but not sufficient condition for the model to have a high predictive power [42]. Therefore, in addition to this statistic value, we also used an external prediction test set. This type of model validation is very important, if we take into account that the predictive ability of a QSAR model can be estimated using only an external test set of compounds (in the model range), which was not used to build the model itself [43].

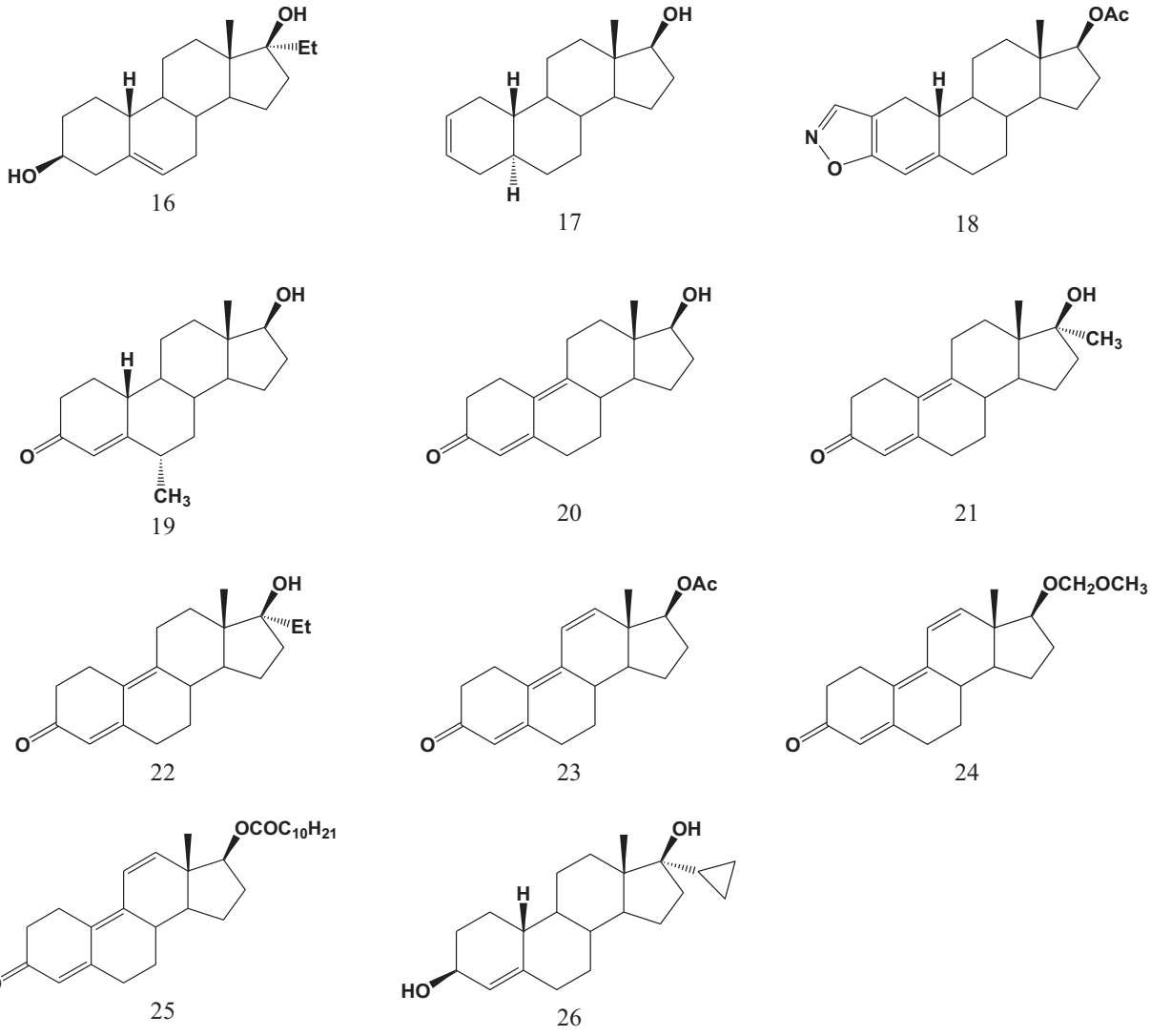


Fig. 2. (*Continued*).

2.4. Clustering

One of the important applications in population study is cluster analysis (CA) based on similarity measurements. The clustering problem can be described as finding "natural groupings" in a set of data. This question actually involves two separate issues: how to measure the similarity between samples, and how to divide sets of samples into clusters among a large number of data input. The term CA currently encompasses a number of different classification algorithms. Most popular algorithms are *k*-Mean cluster algorithms (k-MCA) and Jarvis-Patrick (also known as *k*-nearest neighbor cluster algorithm; *k*-NNCA) algorithms. The *k*-MCA use an exchange

Table 3
Correlation between quantum and physicochemical molecular descriptors included in the QSAR models.

	ΔE_{L-H}[a]	q_6	q_7	q_9	q_{11}	q_{12}	q_{13}	U[b]
ΔE_{L-H}[a]	1	0.09	0.20	0.39	0.08	0.33	0.04	0.06
q_6		1	0.33	0.07	0.00	0.00	0.01	0.25
q_7			1	0.36	0.03	0.05	0.01	0.02
q_9				1	0.06	0.48	0.05	0.01
q_{11}					1	0.00	0.00	0.02
q_{12}						1	0.23	0.05
q_{13}							1	0.15
U								1

[a] Energy difference between the *LUMO* and the *HOMO*. *q*, net atomic charges of carbons 6, 7, 9, 11, 12 and 13, respectively in the steroids backbone.
[b] Chemical potential.

method to divide n data points into k groups (clusters) so that the sum of distances/dissimilarities among the objects within the same cluster is minimized. The k-mean approach requires knowing previously the number of clusters (k). In k-NNCA method, the user must specify the number of nearest neighbors, and the number of neighbors in common to merge two objects [44]. In order to design training and test series and to demonstrate the structural diversity of the present database, we carried out one of these kinds of cluster analyses (k-NNCA) for steroid series. The STATISTICA vs. 5.5, software package [45] was used to develop the CA.

When the clusters are formed the distance among objects (molecules) can be measured by different types of distances. The most straightforward way for computing distances between objects in multidimensional space is to compute Euclidean distances. They are computed as:

$$d_{ij} = \sqrt{\left\{\sum_{k-1}^{p}(x_{ik} - x_{jk})^2\right\}} \quad (2)$$

where x_{ik} is the value of variable x_k for individual i and x_{jk} is the value of some variable for individual j [46].

In addition, there are possibilities of computing various types of distance measure as: square Euclidean distances, Metropolis city-block distances, Chebychev distance and power distances. In our work, we compute squared Euclidean distances. When each object (molecule) represents its own cluster, the distance between those objects are defined by the chosen distance measure. However, when several objects have been linked together, we have to determine the distance between those new clusters. In other words we need a linkage or amalgamation rule to determine when two clusters are sufficiently similar to be linked together. There are numerous linkage rules that have been proposed: single linkage, complete linkage, unweighted pair-group average and weighted pair-group average [46].

3. Results and discussion

3.1. Construction of training and test sets using hierarchical cluster analyses

It is well known that the quality of a regression model is highly dependent on the quality of the selected data set. The most critical aspect for constructing the training set is to warrant enough molecular diversity on it. Taking this into account, we selected a data set of 26 steroids (whit included both anabolic and androgenic activities) having a great structural variability. In order to demonstrate the structural diversity of this data set, we performed a hierarchical CA of these chemicals [47].

The hierarchical clustering approach finds a hierarchy of objects represented by a number of MDs. The dendrogram given in Fig. 3, using the Euclidean distance (X-axis) and the complete linkage (Y-axis), illustrates the results of the k-NNCA developed in this set. As it can be seen in the dendrogram, there are a great number of different subsets, which prove the molecular variability of the selected chemicals in these databases.

It should be remarked that, recently, several authors have developed a classification of steroids using CA [47]. However, this analysis has been presented only for the benchmark steroids with the corresponding globulin affinity. A complete discussion of the clustering is out of the context of the present study, but several interesting features should be noted. The compounds: [**1–4**], [**6–14**] and [**19–25**] are 19-nor-steroids with a carbonyl group (–C=O) in the atom 3 and a vinyl group (–C=C–) in the atoms 4 and 5 (Fig. 2). A few relative easy instances, where the similarity is relatively obvious, may be identified from direct inspection of the molecular structures. The dendrogram should reflect this high logic-visual similarity. A very obvious case lies in molecules **1**, **3** and **4**, contained in the cluster 1. These have very similar physicochemical and quantum chemical properties and, in fact, it is identified in the dendrogram. A similar case exists between molecules **20**, **21** and **22** with a 4, 9, 10-dien-3-one group and the difference is the 17β-hydroxyl, 17α-methyl and 17α-ethyl groups, respectively. This can be early detected in the cluster 2 from simple inspection of Fig. 3. Compounds **11** and **12** also showed high similarity, where the only difference being a single 7α-methyl group in the compound **12**. The compounds **15**, **16** and **26** with a hydroxyl group (–OH) in the atom 3 are contained in the cluster 3 next to the compound **18**.

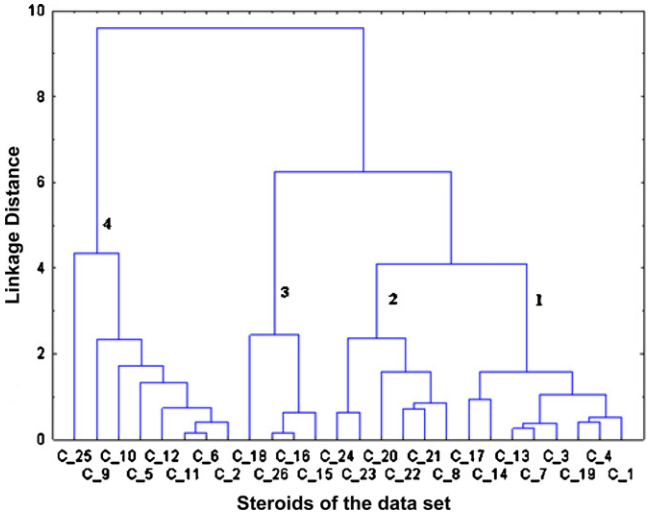


Fig. 3. A dendrogram illustrating the results of the hierarchical k-NNCA of the set of 26 steroids used in the training and prediction sets of the present work.

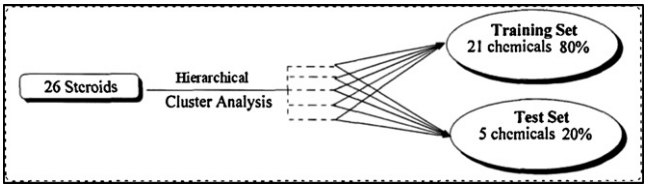


Fig. 4. General algorithm used to design training and test sets throughout k-NNCA.

Furthermore, this procedure allows selecting compounds for the training and test sets, in a representative way, in all level of the linking distance. The main idea of this procedure comprises making a partition of chemicals into several statistically representative classes of compounds. This procedure ensures that any chemical class (as determined by the clusters) will be represented in both compound series. This rational design of training and predicting series allowed us to design both sets that are representative of the whole "experimental universe". Moreover, the selection of the training and prediction sets was performed by taking, in a random way, compounds belonging to each cluster. From these 26 steroids, 21 (80% of the data) were chosen at random to form the training set. The great structural variability of the selected training set makes possible the discovery of leading compounds. The remaining steroids composed of 6 molecules (20% of the data) were prepared as a test set for the external cross-validation of the models. These chemicals have not been used in the development of the QSAR models. Fig. 4 illustrates the above-described procedure in which CA was performed to select a representative sample for the training and test sets.

Table 4
Calculated values and residuals from the anabolic activity relate with the levator ani muscle weight.

Compound[a]	log (1/LA)$_{Cal.}$[b]	Residual[c]
1	1.69	0.41
2	1.74	0.20
3	1.85	0.11
4*	1.88	−0.28
5*	1.86	0.44
6	2.37	−0.15
7	2.48	0.02
8	2.09	−0.24
9	2.11	−0.27
10*	1.98	−0.09
11	1.39	0.37
12	2.03	0.06
13	1.61	−0.01
14	1.37	−0.22
15*	2.63	0.22
16	2.78	−0.18
17	1.50	−0.20
18	2.11	−0.20
19	1.61	−0.20
20	1.89	0.11
21	2.05	0.06
22	2.18	−0.18
23	2.77	−0.07
24*	3.02	0.28
25	2.62	0.08
26	1.84	−0.14

Chemicals marked with asterisk are the steroids of the test set.
[a] Number of compounds given in Table 1 and Fig. 2.
[b] Values of anabolic activity calculated by Eq. (3).
[c] Residual = log (1/LA)$_{Exp.}$ − log (1/LA)$_{Cal.}$

Table 5
Calculated values and residuals from andorgenic activity relate with the ventral prostate weight.

Compound[a]	log (1/VP)$_{Cal.}$[b]	Residual[c]
1	1.18	0.52
2	1.19	0.61
3	1.16	0.76
4*	1.13	−0.66
5*	1.52	0.08
6	2.18	0.04
7	2.17	0.17
8	1.12	0.28
9	1.11	0.04
10*	1.14	−0.66
11	1.15	0.51
12	2.15	−0.07
13	1.14	−0.14
14	0.52	0.33
15*	1.56	0.14
16	1.58	−0.28
17	0.53	−0.23
18	1.27	−0.27
19	1.46	−0.23
20	1.35	−0.35
21	1.34	0.14
22	1.34	−0.34
23	2.79	−0.09
24*	2.90	0.40
25	2.91	−0.21
26	1.38	−0.61

Chemicals marked with asterisk are the steroids of the test set.
[a] Number of compounds given in Table 1 and Fig. 2.
[b] Values of androgenic activity calculated by Eq. 4.
[c] Residual = log (1/VP)$_{Exp.}$ − log (1/VP)$_{Cal.}$.

3.2. Development and validation of the QSAR models

The training set of 19-nor-testosterone derivatives includes a set of compounds formed by the steroids: [**1–3**], [**6–9**], [**11–14**], [**16–24**] and **26** (n = 21, see Fig. 2 and Table 1 for more details). Eq. (3) shows both the variables set from the GA and the best model of anabolic activity. In order to compare the external predictions corresponding to Eq. (3), the steroids: **4**, **5**, **10**, **15**, and **25** were chosen as test set (n = 5, see Fig. 2 and Table 1 for more details). The obtained QSAR model is given below together with the statistical parameters of both learning and prediction sets.

$$\log\left(\frac{1}{\mathrm{LA}}\right) = -0.28(\pm0.13)\Delta E_{L-H} + 25.21(\pm7.94)q_7 - 16.55(\pm11.93)q_{12} + 5.56(\pm2.26) \quad (3)$$

n = 21, R^2 = 0.83, q^2 = 0.77, s = 0.19, F = 27.66, p < 0.001.
Test set: n = 5, R^2 = 0.83, s = 0.30, F = 13.70.

The R^2 indicates that the model explains 83% of the variance for the experimental values of log (1/LA). The model has q^2 = 0.77. This value (q^2 > 0.5) can be considered as a proof of the high predictive ability of the model as well as the good prediction of the test set (R^2 = 0.83). Table 4 shows the correlation between the observed and predicted anabolic activities from Eq. (3).

The VP and SV training set of 19-nor-testosterone derivatives consists of the following chemicals: [**1–3**], [**6–9**], [**11–14**], [**16–24**] and **26** (n = 21, see Fig. 2 and Table 1 for more details). Eqs. (4) and (5) show both, the variables set from the GA and the best models of androgenic activity. In order to compare the predictive ability of VP and SV steroid-based models we used a test set composed of nine compounds **4**, **5**, **10**, **15** and **25**, (see Fig. 2 and Table 1 for more details). The QSAR models obtained for description of VP and SV, as well as their statistical parameters of both training and test sets, are depicted below as Eqs. (4) and (5), respectively:

$$\log\left(\frac{1}{\mathrm{VP}}\right) = 14.11(\pm8.41)q_6 + 54.66(\pm17.53)q_7 + 55.80(\pm19.01)q_9 + 11.34(\pm2.80) \quad (4)$$

n = 21, R^2 = 0.76, q^2 = 0.65, s = 0.34, F = 18.30, p < 0.001.
Test set: n = 5, R^2 = 0.86, s = 0.47, F = 21.10.

$$\log\left(\frac{1}{\mathrm{VS}}\right) = 14.50(\pm8.79)\,q_6 + 55.72(\pm18.32)q_7 + 54.85(\pm19.87)q_9 + 11.39(\pm2.93) \quad (5)$$

n = 21, R^2 = 0.74, q^2 = 0.61, s = 0.34, F = 16.91, p < 0.001.
Test set: n = 5, R^2 = 0.92, s = 0.36, F = 31.92.The R^2 for Eqs. (4) and (5) are 0.76 and 0.74, respectively. Therefore, these models explained 76% and 74% of the variance for the experimental values of log (1/VP) and log (1/SV). These models also showed high predictive ability (q^2 of 0.65 and 0.61, and R^2 of 0.86 and 0.92, respectively). Tables 5 and 6 show the correlation between observed and predicted values of androgenic activities for Eqs. (4) and (5), respectively.

In order to design compounds which preserve a high degree of anabolic activity and a vastly diminished androgenic activity, AAR values: log (1/LA)/log (1/VP) and log (1/LA)/log (1/SV) of 19-nor-testosterone steroids were estimated. The AAR values were quantified using the anabolic and androgenic activities values shown in Table 1. The training set of (log (1/LA)/log (1/VP)) ratio includes a set of compounds formed by the steroids: **1**, **3**, [**6–11**], **13**, **14**, **16**, [**18–23**], **25** and **26** (n = 19, see Fig. 2 and Table 1). The statistical outliers: 17α-ethynyl-19-nor-testosterone (**4**), 7α-methylestr-4-en-3, 17-dione (**12**) and 17β-hydroxy-5α-estr-2-ene (**17**) (n = 3, see Fig. 1) were removed from the database. The training set of (log (1/LA)/log (1/SV)) ratio includes a set of compounds formed by the steroids: **1**, **3**, **4**, [**6–9**], [**11–14**], **16**, [**18–23**], **25** and **26** (n = 20, see Fig. 2 and Table 1). The statistical outliers: 4-chloro-

Table 6
Calculated values and residuals from anabolic activity relate with the seminal vesicle weight.

Compound[a]	log (1/SV)$_{Cal.}$[b]	Residual[c]
1	1.19	0.24
2	1.20	0.55
3	1.18	0.56
4*	1.15	−0.15
5*	1.56	0.05
6	2.26	−0.10
7	2.26	0.39
8	1.13	0.41
9	1.12	0.14
10*	1.15	−0.67
11	1.16	0.04
12	2.24	−0.06
13	1.16	−0.16
14	0.50	0.35
15*	1.59	0.10
16	1.61	−0.31
17	0.51	−0.21
18	1.28	0.02
19	1.49	−0.26
20	1.35	−0.35
21	1.34	0.14
22	1.35	−0.35
23	2.80	−0.10
24*	2.91	0.39
25	2.92	−0.22
26	1.41	−0.63

Chemicals marked with asterisk are the steroids of the test set.
[a] Number of compounds given in Table 1 and Fig. 2.
[b] Values of androgenic activity calculated by Eq. 5.
[c] Residual = log (1/SV)$_{Exp.}$ − log (1/SV)$_{Cal.}$.

Table 7
log (1/LA)/log (1/VP) ratio: experimental, calculated values and residuals from Eq.6.

Compound[a]	log(1/LA)/log (1/VP)$_{Exp.}$[b]	log(1/LA)/log (1/VP)$_{Cal.}$[c]	Residual[d]
1	1.23	1.12	0.11
2*	1.08	1.53	−0.45
3	1.01	0.93	0.08
4	3.36	1.47	1.89
5*	1.44	2.62	−1.19
6	1.00	0.80	0.20
7	1.07	0.99	0.08
8	1.32	1.44	−0.12
9	1.61	1.51	0.10
10	1.07	1.26	−0.19
11	1.01	1.24	−0.24
12	3.95	1.66	2.29
13	1.60	1.33	0.27
14	1.36	1.43	−0.07
15*	1.67	2.27	−0.59
16	2.00	2.12	−0.12
17	4.32	2.51	1.81
18	1.90	1.58	0.33
19	1.15	1.18	−0.03
20	2.00	1.91	0.09
21	1.43	1.71	−0.27
22	2.00	1.70	0.30
23	1.00	1.30	−0.30
24*	1.00	1.65	−0.65
25	1.00	1.40	−0.40
26	2.18	1.99	0.20

Chemicals marked with asterisk are the steroids of the test set.
[a] Number of compounds given in Table 1 and Fig. 2.
[b] Experimental values of the effective log (1/LA)/log (1/VP) ratio.
[c] Values calculated by Eq. (6).
[d] Residual = log(1/LA)/log (1/VP)$_{Exp.}$ − log(1/LA)/log(1/VP)$_{Cal.}$.

17β-acetoxyestr-4-en-3-one (**10**) and 17β-hydroxy-5α-estr-2-ene (**17**) (n = 2, see Fig. 2) were removed from the database. Outlier detection was carried out using the following standard statistical tests: residual, standardized residual and Cooks distance. The QSAR models obtained for the AAR apparently do not describe the stereo-electronic effect of these molecules.

In order to compare the external predictions corresponding to Eqs. (6) and (7), the steroids: **2**, **5**, **15**, and **24** were chosen as test set (n = 4, see Fig. 1 and Table 1 for more details). The MDs selected by the AG are shown in Eqs. (6) and (7):

$$\frac{\log(1/\mathrm{LA})}{\log(1/\mathrm{VP})} = 0.41(\pm0.33)\mu - 16.07(\pm11.50)\,q_{13} + 2.08(\pm0.78)U + 9.27(\pm2.75) \quad (6)$$

n = 19, R^2 = 0.70, q^2 = 0.54, s = 0.24, F = 12.47, p < 0.001.
Test set: n = 4, R^2 = 0.66, s = 0.22, F = 8.60.

$$\frac{\log(1/\mathrm{LA})}{\log(1/\mathrm{SV})} = 9.74(\pm8.69)q_{11} + 0.84(\pm0.53)U + 6.58(\pm2.53) \quad (7)$$

n = 20, R = 0.58, q = 0.35, s = 0.27, F = 11.30, p < 0.001.
Test set: n = 4, R^2 = 0.72, s = 0.19, F = 11.44.

The best regression QSAR model for any AAR was obtained by Eq. (6). The R^2 indicates that the model explains 70% of the variance for the experimental values of log (1/LA)/log (1/VP) ratio and this model has an adequate q^2 = 0.54. This value (q^2 > 0.5) and R^2 of the test set (R^2 = 0.70) can be considered as a proof of the high predictive ability of the model. On the other hand, Eq. (7) shows an adequate fitness (R^2 = 0.58) but very low predictive power (q^2 = 0.35). This value of q^2 < 0.5 can be considered as a proof of the low predictive ability of the model.

Table 7 shows experimental and calculated log (1/LA)/log (1/VP) ratio values as well as residuals from the best regression model (Eq. (6)). The model predicts that the 17α-cyclopropyl-17β, 3β-hydroxy-4-estrene (**26**) compound present the highest anabolic–androgenic ratio (AAR) and the 7α-methyl-17β-acetoxy-estr-4-en-3-one (**6**) the lowest AAR.

3.3. Driving forces for biological activities of 19-nor-testosterone Steroids

Interrelations of MDs make difficult the interpretation of the QSAR model. Therefore, it is well known that the interrelatedness among the different MDs results in highly unstable correlation coefficients, which makes it impossible to know the relative importance of an index and underestimates the utility of the regression coefficient in a model [48]. However, in some cases strongly interrelated descriptors can enhance the quality of a model, because the small fraction of a descriptor that is not reproduced by its strongly interrelated pair can provide positive contributions to the modeling. On the other hand, the coefficients of the QSAR model based on orthogonal descriptors are stable to the inclusion of novel descriptors, which permit to interpret the regression coefficients and to evaluate the role of individual molecular fingerprints in the QSAR model. Calculated quantum and physicochemical molecular descriptors were subjected to an intercorrelation study (see Table 3). Correlation between variables included in each QSAR model was rather low, indicating the different information contents of each term in these equations.

The QSAR model of Eq. (3) has three types of electronic molecular descriptors: ΔE_{L-H}, q_7 and q_{12}. These three properties are most responsible for the anabolic activity expressed by log (1/LA). Negative ΔE_{L-H} terms indicate that the high values of ΔE_{L-H} produce unfavorable anabolic effects. It is observed that the presence of a 4, 9, 11-triene group (compounds **23**, **24** and **25**, Fig. 5) decreases the ΔE_{L-H} value and increases the anabolic activity (Table 2). The negative coefficient of q_{12} descriptor indicates that negative charge in C-12 (ring C) increases the activity of these compounds (see Eq. (3)). Moreover, an electron-donating substituent in the C-7 produces an

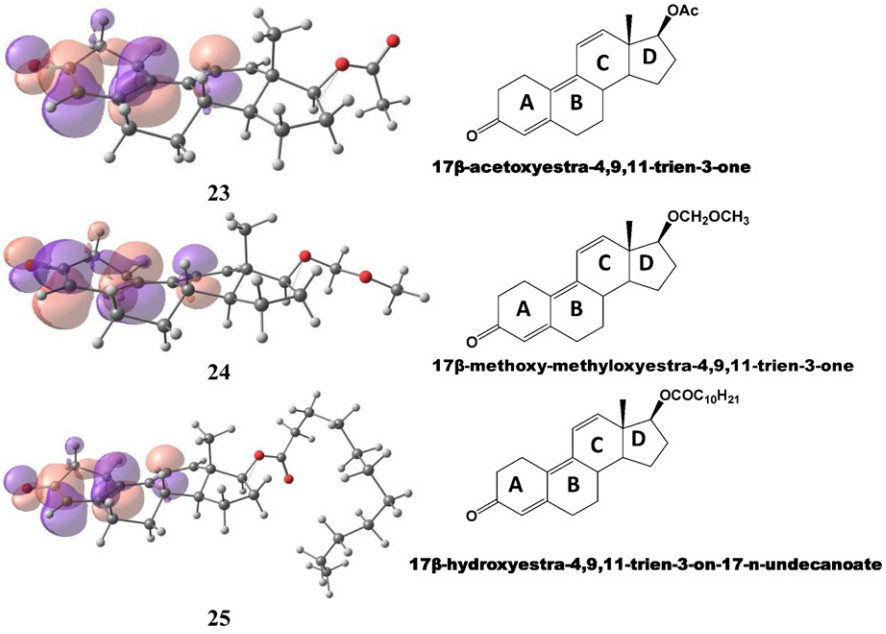


Fig. 5. Graphic representation of the energy difference between the *LUMO* and the *HOMO* (ΔE_{L-H}) for the most active compounds: **23**, **24** and **25**.

unfavorable anabolic effect. The charge value of atom C-7 is largely dependent on the substitution pattern of ring B. It is observed that the presence of groups 7α-methyl (compounds **6**, **7**, and **12**) and 5-ene (compounds **15** and **16**) in the ring B increases the positive charge value on C-7 and increases the binding affinity for anabolic receptor.

Models for androgenic activity (Eqs. (4) and (5)) also explain the steroid–receptor interaction. It is mostly due to the biological activity expressed by electronic descriptors (q_6, q_7 and q_9). Again, the charge value of C-7 has positive contribution to the androgenic activity. These equations explain the increase in the androgenic activity related to the increase in the positive charge of atoms 6 (ring B) and 9 (fusion point between rings B and C). Therefore, positive q_6, q_7 and q_9 values indicate that electron-withdrawing substituent in the carbons 6, 7 and 9 correlated with a higher androgenic activity.

Finally, Eq. (6) shows that the AAR is favorable with the increase in the negative charge of atom 13 (fusion point between rings C and D), high values of μ and *U*. The charge in atom 13 should mostly be influenced by the substituent in fusion point between rings C and D.

3.4. The applicability domain of the QSAR models

A crucial problem in chemometric and QSAR studies is the definition of the applicability domain (AD) of a classification or regression model. Even robust, significant and validated QSAR model will not be used in order to predict the modeled property for the entire universe of chemicals. In fact, only the predictions for chemicals falling within this domain can be considered reliable and not model extrapolations [49]. The AD is a theoretical region in chemical space, defined by the model descriptors and modeled response, and thus by the nature of the chemicals in the training set, as represented in each model by specific MDs. Moreover, AD of the QSAR model is the range within which it tolerates a new molecule [49].

It is generally acknowledged that QSAR models are only valid within the same domain for which they were developed. In fact, even when the models are developed on the same chemicals, the AD for new chemicals can differ from model to model, depending on the specific descriptors. However, model validation is sometimes neglected and the application domain is not always well defined [49]. The purpose of this section is to outline how validation and domain definition determine in which situation it is correct to use the model. The aim of the present work was to develop a model for predicting AAR of steroids at early stages of drug discovery and development. Accordingly, we only selected 19-nor-testosterone analogues. Consequently, it cannot extrapolate the use of these models to other kinds of class-steroids making uncertain prediction in conditions that are very different to those fixed for derive the model [50]. It is important to note that in multiple predictor models, simple single-variable range checks are not sufficient to verify AD. At present, there are several approaches to evaluate the AD of QSAR models. For RLM, a multiple predictor problem with normally distributed data, the distance-based measure, as leverage is one of most used. Through the leverage approach [51] it is possible to verify whether a new chemical will lie within the structural model domain. The leverage *h* [52] of a compound measures its influence on the model. Namely, leverage used as a quantitative measure of the model AD is suitable for evaluating the degree of extrapolation, which represents a sort of compound distance from the experimental model space. Leverage values can be calculated for both training compounds and new compounds. In the first case, they are useful to find training compounds that influence model parameters to a marked extent, resulting in an unstable model. In the second case, they are useful to check the applicability domain of the model [49]. The warning leverage, h^*, is a critical value or cut-off to consider the prediction made for the model for specific compounds in data set. The leverage h^* can be defined as $3 \times p'/n$, where n is the number of training chemicals and p' is the number of model parameters plus one [49]. Prediction should be considered unreliable for compounds of high leverage value ($h > h^*$). A leverage greater than the warning leverage h^* means that the compound-predicted response can be extrapolated from the model, and therefore, the predicted value must be used carefully. Only predicted data for chemicals belonging to the chemical domain of the training set should be proposed.

The AD of a QSAR model is visualized by the Williams plot, a double-ordinate Cartesian plot of standardized residuals (*Y*-axis) versus leverage (hat diagonal; *X*-axis) values for each compound of the training and test set. From this plot the AD is established inside

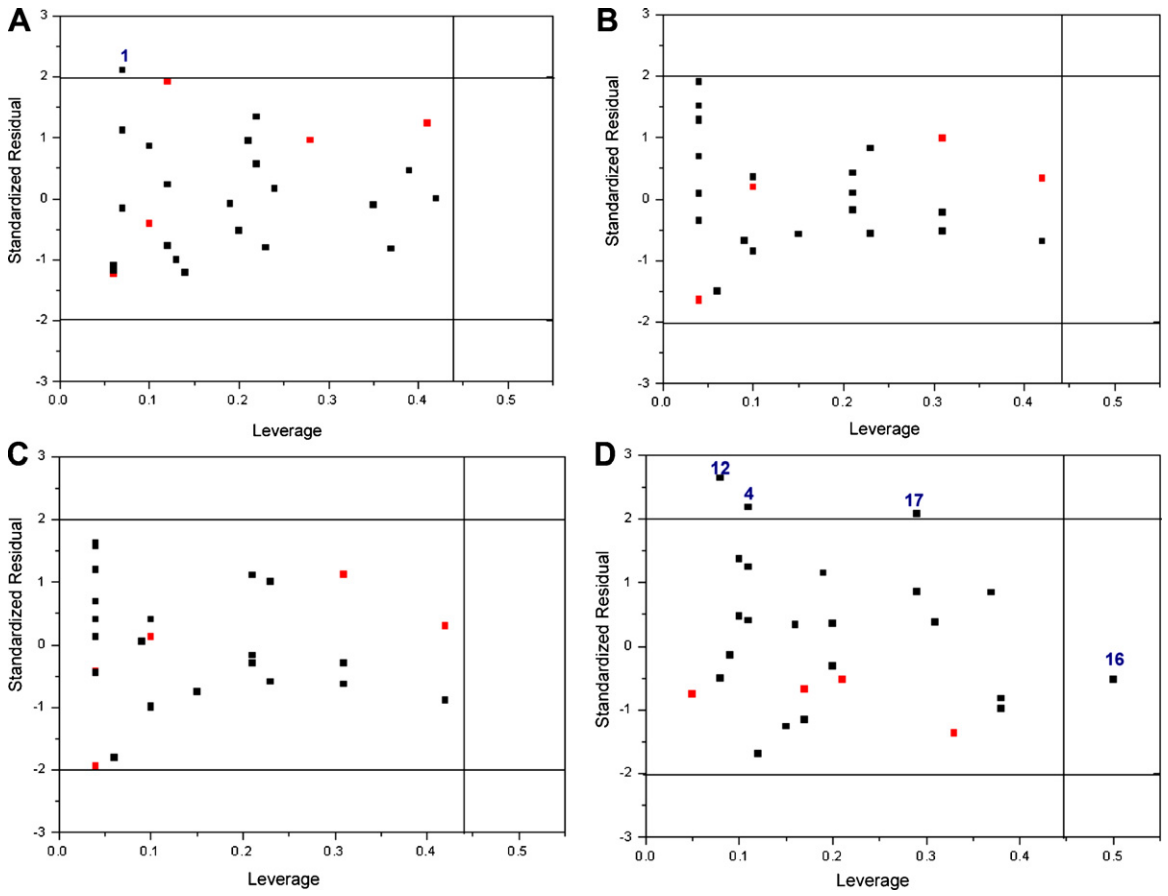


Fig. 6. William plots of Eqs. (3)–(6) are shown in (A)–(D), respectively. Black and red points represent the steroids of the training and test series, respectively. Outlier compounds [**4**, **12** and **17**] are points with standardized residual greater than two standard deviation unit. Influential chemical [**16**] is point with high leverage values higher than the threshold or cut-off value $h = 0.44$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

a square area within ±2. SIGMA (SIGMA, standard deviation) band for residuals and a leverage threshold h^*. The Williams plot can be used for an immediate and simple graphical detection of both the response outliers (i.e., compounds with standardized residual greater than two standard deviation unit, $>2\sigma$) and structurally influential chemicals in a model ($h > h^*$). For instance, Fig. 6(A–D) shows William plots of Eqs. (3)–(6), respectively. As noted from Fig. 6, all used steroids lie within this area. In Fig. 6A, the compound **1** has higher standardized residuals than the threshold but shows leverage within the limits. The structure of this steroid is very close to other compounds of the training set. Outliers compounds (**4**, **12** and **17**) were detected in Fig. 6D. These compounds slightly exceed the critical hat values of standardized residual greater than two. Influential chemical [16] point with leverage values higher than the threshold or cut-off value $h = 0.44$ is also detected. However, this model can be used with high accuracy in these AD.

4. Conclusions

In the present report, predictive QSAR models of the anabolic and androgenic activities of the 19-nor-testosterone steroids family were carried out by employing quantum and physicochemical MDs, as well as, a GA for the selection of variables. The MDs included in the reported models allow the structural interpretation of the biological process, making evident the main role of the electronic properties in this steroid family. The selected QSAR equation for anabolic and androgenic activities explains the steroid–receptor interaction mainly due to the biological activities expressed by the electrostatic properties.

Three electronic descriptors: ΔE_{L-H}, q_7 and q_{12} are found mostly responsible for the anabolic activity, of these compounds, expressed by log (1/LA). Negative ΔE_{L-H} terms indicate that the high values of ΔE_{L-H} produce unfavorable anabolic effects. It is observed, that the presence of a 4, 9, 11-triene group (compounds **23**, **24** and **25**) decreases the ΔE_{L-H} value and increases the anabolic activity. The negative coefficient of q_{12} descriptor indicates that negative charge in C-12 of ring C increases the activity of these compounds (see Eq. (2)). Moreover, an electron-donating substituent in the C-7 produces an unfavorable anabolic effect. The charge value of atom C-7 is largely dependent on the substitution pattern of ring B. The presence of groups 7α-methyl (compounds **6**, **7**, and **12**) and 5-ene (compounds **15** and **16**) in ring B increases the positive charge value on C-7 and the binding affinity for anabolic receptor. Values of ΔE_{L-H} show a correlation better than 0.7. For this reason, we have selected this descriptor and none other descriptors related with this value: chemical hardness (η), chemical softness (S) and electrophilicity index (ω). However, our results are according with that obtained by Singh et al. [53,54] in QSAR studies on testosterone and derivatives.

Models for androgenic activity also explain the steroid–receptor interaction. It is mostly due to the biological activity expressed by electronic descriptors (q_6, q_7 and q_9). Again, the charge value of C-7 has positive contribution to the androgenic activity. The increasing in the androgenic activity is related to the increasing in the positive charge of C-6 (ring B) and C-9 (fusion point between rings B and C). Therefore, positive q_6, q_7 and q_9 values indicate that electron-withdrawing substituent in the carbons 6, 7 and 9 are correlated with a higher androgenic activity.

For increasing the negative charge of C-13 (fusion point between rings C and D), high value of molecular dipole moment and high value of chemical potential can explain a favorable AAR for these steroids. The charge on C-13 should be mostly influenced by the substituent in fusion point between rings C and D. Calculated values for the AAR predict that the 17α-cyclopropyl-17β, 3β-hydroxy-4-estrene (**26**) compound present the highest AAR and the 7α-methyl-17β-acetoxy-estr-4-en-3-one (**6**) the lowest AAR in agreement with the experimental data. Finally, the AD of the developed models was assessed and visualized by Williams plots as a squared area within ±2.(standard deviation) band for residuals and a leverage threshold of $h = 0.44$.

Acknowledgments

Most of the authors are indebted to the Universidad de La Habana, Cuba, for the overall support of their education and work. The Universidad Autónoma de Madrid served as excellent host for the essential part of this work as well as the Spanish Agency for International Cooperation for Development (D/019558/08 and D/023653/09) funding. JMGV is grateful for support from MEC of Spain (CTQ 2007-63332 and CTQ2010-19232).

References

[1] T. Ingemar, P. Anna, Fundam. Clin. Pharmacol. 19 (2004) 27.
[2] M. Montano, J.N. Flanagan, L. Jiang, P. Sebastiani, M. Rarick, N.K. LeBrasseur, C.A. Morris, R. Jasuja, S. Bhasin, J. Clin. Endocr. Metab. 92 (2007) 2793.
[3] G.S. Lynch, J.D. Schertzer, J.G. Ryall, Pharmacol. Ther. 113 (2007) 461.
[4] I. Stoian, B. Manolescu, V. Atanasiu, O. Lupescu, Cent. Eur. J. Med. 2 (2007) 361.
[5] F. Cosman, Curr. Osteoporos. Rep. 8 (2006) 63.
[6] M. Bossola, F. Pacelli, A. Tortorelli, G.B. Doglietto, Ann. Surg. Oncol. 14 (2007) 276.
[7] J.D. Wilson, Endocr. Rev. 9 (1998) 181.
[8] S.J. Winters, in: G.C. Lin, L. Erinoff (Eds.), Anabolic Steroid Abuse Androgens, NIDA Research Monographs Series US Department of Health and Human Services, Rockville, MD, 1990, pp. 131–141.
[9] W. Llewellyn, Anabolics 2007: anabolic steroids reference manual, Body Sci. (2007).
[10] W. Llewellyn, Anabolics 2004, Mol. Nutr. (2003).
[11] F.A. Pasha, H.K. Srivastava, P.P. Singh, Int. J. Quantum Chem. 104 (2005) 87.
[12] R. Kiralj, Y. Takahata, M.M.C. Ferreira, QSAR Comb. Sci. 22 (2003) 430.
[13] D.G. Carolina, B.K. Lemont, H. Lowell, J. Comput. Aided Mol. Des. 12 (1998) 557.
[14] S. Anzali, J. Gasteiger, U. Holzgrabe, J. Polanski, J. Sadowski, A. Teckentrup, M. Wagener, in: H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in Drug Design, vol. 2, Kluwer/ESCOM, Dordrecht, NL, 1998, pp. 273–299.
[15] C.T. Klein, D. Kaiser, G. Ecker, J. Comput. Aided Mol. Des, 44 (2004) 200.
[16] K. Tuppurainen, M. Viisas, M. Peräkylä, R. Laatikainen, J. Chem. Infor. Comput. Sci, 42 (2002) 607.
[17] J. Polanski, A. Bak, R. Gieleciak, T. Magdziarz, Molecules 9 (2004) 1148.
[18] M.A. Klieber, C. Underhill, G.L. Hammond, Y.A. Muller, J. Biol. Chem. 282 (2007) 29594.
[19] Y.M. Alvarez-Ginarte, R. Crespo-Otero, L.A. Montero-Cabrera, J.A. Ruiz-García, Y. Marrero-Ponce, R. Santana, F. Pardillo, E. Alonso, QSAR Comb. Sci. 24 (2005) 218.
[20] Y.M. Alvarez-Ginarte, R. Crespo-Otero, Y. Marrero-Ponce, L.A. Montero-Cabrera, J.A. Ruiz-Garcia, A. Padron, F. Torrens-Zaragoza, QSAR Comb. Sci. 25 (2006) 881.
[21] Y.M. Alvarez-Ginarte, R. Crespo-Otero, Y. Marrero-Ponce, P. Noheda-Marin, J.M. García de la Vega, L.A. Montero-Cabrera, J.A. Ruiz-García, Bioorg. Med. Chem. 16 (2008) 6448.
[22] Y.M. Alvarez-Ginarte, Y. Marrero-Ponce, J.A. Ruiz-Garcia, L.A. Montero-Cabrera, F. Torrens-Zaragoza, R. Garcia-Domenech, J. Comput. Chem. 29 (2008) 317.
[23] J.A. Vida, Androgens and Anabolic Agents. Chemistry and Pharmacology, Academic Press, New York, 1969.
[24] C.G. Wermuth, The Practice of Medicinal Chemistry, Academic Press, London, 1996.
[25] K. Ghose, A. Pritchett, G.M. Crippen, J. Comput. Chem. 9 (1998) 80.
[26] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim (Ger.), 2000.
[27] T. Ooi, M. Oobatake, G. Nemethy, H.A. Scheraga, Proc. Nat. Acad. Sci. U.S.A. 84 (1987) 3086.
[28] K.J. Miller, J. Am. Chem. Soc. 112 (1990) 8533.
[29] R.G. Parr, P.K. Chattaraj, J. Am. Chem. Soc. 113 (1991) 1854.
[30] R.G. Parr, L.V. Szentpály, S. Liu, J. Am. Chem. Soc. 12 (1999) 1922.
[31] J.J.P. Stewart, J. Comput. Chem. 10 (1989) 209.
[32] J.J.P. Stewart, MOPAC, v. 6. 0, QCPE Bull. 10 (1990) 86.
[33] W. Hasel, T.F. Hendrickson, W.C. Still, Tetrahedron Comput. Methodol. 1 (1988) 103.
[34] A.K. Ghose, G.M. Crippen, J. Chem. Inform. Comput. Sci. 27 (1987) 21.
[35] D.B. De Oliveira, A.C. Gaudio, QSAR Comb. Sci. 19 (2003) 599.
[36] K. Hasegawa, T. Kimura, K. Funatsu, Quantum Struct. Act. Relat. 18 (1999) 262.
[37] D.B. Turner, P. Willett, J. Comput-Aided Mol. Des. 14 (2000) 1.
[38] D.A. Coley, Introduction to Genetic Algorithms for Scientists and Engineers, World Scientific, New Jersey River Edge (NJ), 1999.
[39] M.C. Ford, D.C. Salt, Chemometric Methods in Molecular Design, VCH Publishers, Weinhiem (Ger.), 1995.
[40] A. Golbraikh, A. Tropsha, J. Comput. Aided Mol. Des. 16 (2002) 357.
[41] S. Wold, L. Erikson, Statistical Validation of QSAR Results. Validation Tools in Chemometric Methods in Molecular Design, VCH Publishers, Weinhiem (Ger.), 1995.
[42] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269.
[43] G. Schüürmann, R.U. Ebert, J. Chen, B. Wang, R. Kühne, J. Chem. Inf. Model. 48 (2008) 2140.
[44] J.W. Farland, D.J. Gans, Chemometric Methods in Molecular Design, VCH Publishers: Wiley, Weinheim (Ger.), 1995.
[45] StatSoft (1999). STATISTICA for Windows, v. 5.5. Tulsa(OK), USA.
[46] H. Xu, Curr. Top. Med. Chem. 2 (2002) 1305.
[47] G. Restrepoa, J.L. Villaveces, Chem. Acta 78 (2005) 275.
[48] M. Randić, J. Mol. Struct. (Theochem.) 233 (1991) 45.
[49] P. Gramatica, QSAR Comb. Sci. 26 (2007) 694.
[50] E. Papa, F. Villa, P. Gramatica, J. Chem. Inf. Model. 45 (2005) 1256.
[51] A.C. Atkinson, Transformations and Regression, Clarendon Press, Oxford, 1985.
[52] H. Liu, E. Papa, P. Gramatica, Chem. Res. Toxicol. 19 (2006) 1540.
[53] P.P. Singh, H.K. Srivastava, F.A. Pasha, Bioorg. Med. Chem. 12 (2004) 171.
[54] H.K. Srivastava, F.A. Pasha, P.P. Singh, Int. J. Quantum Chem. 103 (2005) 237.