ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2002) 345–358

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Structure–activity relationships (SAR) of contraceptive progestogens studied with four different methods using calculated physicochemical parameters

Rosana Vendrame[1], Márcia M.C. Ferreira, Carol H. Collins, Yuji Takahata*

Universidade Estadual de Campinas, Instituto de Química, Caixa Postal 6154 Campinas, São Paulo 13081-970, Brazil

Received 17 October 2000; received in revised form 1 May 2001; accepted 2 May 2001

Abstract

Structure–activity relationships (SAR) of the contraceptive progestogens for (I) oral contraceptive activity (OCA), (II) androgenic effect, and (III) binding affinity for sex hormone binding globulin (SHBG) were studied using four different methods: principal component analysis (PCA), hierarchical cluster analysis (HCA), neural networks (NN), and electronic indices method (EIM) employing descriptors calculated by the semi-empirical Austin Model 1 (AM1) method. An additional set of molecules was used to check the reliability of the results obtained for OCA by PCA. Using PCA, three different sets of descriptors were found to correlate with the three different biological activities, I–III, indicating that the interaction between the receptor and the progestogen must depend on the type of biological activity. The descriptors selected by PCA were also employed for SAR analysis of the contraceptive progestogens using two other methods, HCA and NN. Both HCA and NN correctly classified high activity molecules as different from low activity ones. Thus, those descriptors selected by PCA work well in the other two methods of classification. Using the sign of ρ, a difference of electron densities of selected molecular orbitals in a specified region in a molecule, it was possible to discriminate high activity molecules from low activity molecules in the three different types of activities studied, I–III, with one exception. © 2002 Elsevier Science Inc. All rights reserved.

Keywords: SAR; Contraceptive progestogens; Calculated physicochemical parameters; AM1; Electronic indices method; Neural network

1. Introduction

Combined oral contraceptives (COCs) have been widely used all over the world for more than 3 decades. The COCs are generally obtained from the combination between ethynylestradiol (EE) and one of three new progestogens: desogestrel (DSG), gestodene (GSD) and norgestimate (NGM). The new generation of COCs are low-dose formulations and highly effective. They are an improvement over older low-dose formulations and are clearly preferable to high-dose ones. Side effects of the new COCs generally occur in less than 6% of the cases, within the number of subjects tested [1]. Though the percentage of the side effects is quite small, there still remains room for improvement. The ideal contraceptive would be highly effective, safe, long acting but readily reversible and virtually free of side effects [2].

Orally active progestogens are not purely progestationals. Most of them exhibit androgenic activity as well as a progestational one. Androgenic activity increases the occurrence of side effects such as acne, hirsutism, increase of weight, alterations in the carbohydrate and lipoprotein metabolism and hypertension. A progestogen can also *indirectly* cause androgenic side effects by binding to sex hormone binding globulin (SHBG). Progestogen stimulates androgenic effects by dislocating testosterone from SHBG and increasing the levels of free active testosterone circulating in blood. The binding of progestogen to SHBG is an important measure of its androgenicity [3].

The aim of the present work is to establish a structure–activity relationship (SAR) for a series of steroids, including three new generation progestogens. The biological activities studied were (I) oral contraceptive activity (OCA), (II) androgenic effect, and (III) binding affinity to SHBG. The molecules included in this study are: (**1**) progesterone (P), (**2**) norethindrone (NET), (**3**) NGM, (**4**) levonorgestrel (LNG), (**5**) DSG, (**6**) GSD, (**7**) 17-deacetylnorgestimate (ANGM), and (**8**) 5α-dihydrotestosterone (DHT) (Fig. 1). The steroids **3**, **5** and **6** are the three new progestogens.

The OCA, on a log 1/IC scale where IC is the daily molar concentration required to inhibit the ovulation, for the five steroids, **2**–**6** are between approximately 6 and 7, whereas

* Corresponding author.
E-mail address: taka@iqm.unicamp.br (Y. Takahata).

[1] Present address: Instituto de Física "Gleb Wataghin", Universidade Estadual de Campinas, Campinas, São Paulo 13083-970, Brazil.

1093-3263/02/$ – see front matter © 2002 Elsevier Science Inc. All rights reserved.
PII: S1093-3263(01)00107-3

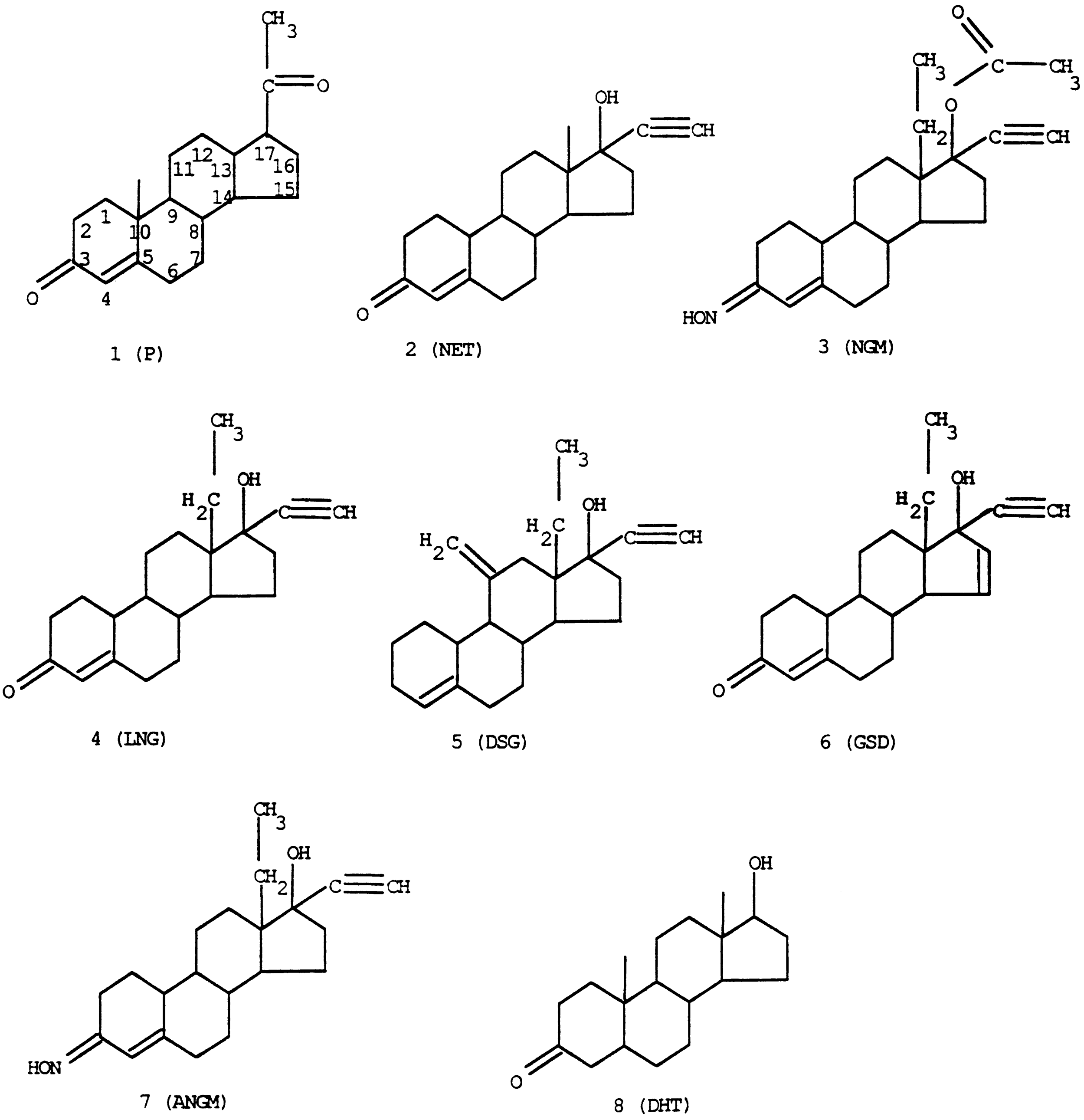

Fig. 1. Progestogens studied: (**1**) P, (**2**) NET, (**3**) NGM, (**4**) LNG, (**5**) DSG, (**6**) GSD, (**7**) ANGM, and (**8**) DHT.

the OCA of **1** is about 3 (Table 1) [4]. We classify the six steroids into two groups: high activity and low activity. Five steroids, **2**–**6**, belong to the high activity group, while progesterone, **1**, belongs to the low activity group. It is interesting to investigate the causes for the large variation of OCA in the two groups. A glance at the six compounds in Fig. 1 indicates that those steroids having ethynyl (–C≡CH) at the 17α and hydroxyl (–OH) or acetoxyl [CH_3COO–] at the 17β positions show high OCA, as in **2**–**6**, whereas those compounds having neither of these at position 17, as is the case for **1**, show low OCA. We want to study the causes of the large OCA variation for the steroids, based upon their calculated physicochemical parameters. We first establish a relationship between the OCAs and the physicochemical parameters. This type of study is commonly designated as SAR. We aim to understand the causes of the difference in steroid OCA mainly at a molecular level.

The androgenic effect of the same group of molecules is also available in the literature [4] and it is listed in Table 2. Molecule **4** shows the highest androgenic effect of this molecular series. The value of the androgenic effect of molecule **4** (LNG) is 40, while for the others it is less than 5. Why does molecule **4** show an androgenic effect more than 10 times greater than the rest of the group? What sort

Table 1
Oral contraceptive activity [4] in log 1/IC and the five physicochemical descriptors selected with PCA[a]

Molecule	log 1/IC[b]	PCA					EIM	
		Q_{10}	Q_{13}	Q_{17}	I	η	ρ	Δ
1 (P)	3.02	-1.89×10^{-2}	-3.90×10^{-2}	-1.54×10^{-1}	10.055	5.002	−0.0028	0.2660
2 (NET)	5.77	-7.89×10^{-2}	-7.20×10^{-2}	1.90×10^{-1}	10.001	4.985	0.0296	0.3184
3 (NGM)	6.17	-7.04×10^{-2}	-4.10×10^{-2}	2.24×10^{-1}	9.050	4.652	0.0053	1.3624
4 (LNG)	6.62	-7.74×10^{-2}	-6.20×10^{-2}	1.94×10^{-1}	10.012	4.988	0.0293	0.3193
5 (DSG)	6.71	-5.77×10^{-2}	-6.10×10^{-2}	1.94×10^{-1}	9.220	5.210	0.0428	0.4009
6 (GSD)	6.89	-7.72×10^{-2}	-6.60×10^{-2}	2.16×10^{-1}	9.917	4.935	−0.0615	0.2282
7 (ANGM)		-7.11×10^{-2}	-6.00×10^{-2}	1.98×10^{-1}	9.097	4.664	0.0034	1.3958
8 (DHT)		-3.67×10^{-2}	-4.10×10^{-2}	0.31×10^{-1}	10.215	5.578	−0.0073	0.1281

[a] Net atomic charges (Q_n), ionization potential (I in eV) and hardness (η in eV) for the progestogen contraceptives and related compounds. The OCA for **7** and **8** is not known. The last two columns list descriptors, ρ and Δ, defined by Eqs. (1) and (2) that are used by the EIM method.

[b] The original data in [4] were given as inhibition of ovulation (mg per day). The data were converted to IC, the molar concentration of the drug necessary daily to inhibit ovulation.

Table 2
Androgenic effect [4] and three selected parameters: frontier radical density in position 7 ($F_7^{(r)}$) of SS, frontier electron density in position 9 ($F_9^{(e)}$) of SS and frontier radical density in position 9 ($F_9^{(r)}$) of SS of progestogen contraceptives[a]

Molecule	Androgenic effect	PCA			EIM
		$F_7^{(r)}$	$F_9^{(e)}$	$F_9^{(r)}$	ρ'
4 (LNG)	+40	2.85×10^{-2}	5.10×10^{-2}	3.36×10^{-2}	0.0007
2 (NET)	+2.7	2.26×10^{-2}	4.24×10^{-2}	1.67×10^{-2}	−0.0002
1 (P)	+1.3	2.65×10^{-2}	3.08×10^{-2}	1.98×10^{-2}	−0.0039
3 (NGM)	+1.0	1.02×10^{-2}	1.45×10^{-2}	1.19×10^{-2}	−0.2207
5 (DSG)	<1.0	2.03×10^{-2}	2.63×10^{-2}	2.14×10^{-2}	−0.0149
6 (GSD)	<1.0	7.80×10^{-3}	4.53×10^{-3}	8.87×10^{-3}	1.3347
7 (ANGM)		1.56×10^{-2}	1.86×10^{-2}	1.56×10^{-2}	−0.2180
8 (DHT)		1.66×10^{-2}	3.00×10^{-2}	1.70×10^{-2}	−0.0698

[a] Androgenic effect for **7** and **8** is not known. The last column lists a descriptor, ρ', defined by Eq. (1) that is used by the EIM method.

of molecular properties are related with this fact? Again, we classify the whole group of molecules into two categories: high activity and low activity. Molecule **4** belongs to the high activity classification, and the molecules **1–3**, **5** and **6** belong to the low activity classification. We shall look for physicochemical parameters that are related to the difference in the androgenic effect of these progestogens.

Values of relative binding affinities for SHBG were taken from the literature [5] and are reproduced in Table 3. Molecule **8**, DHT, is the reference molecule, exhibiting a binding affinity of 100% to SHBG [5]. The molecules **6**, **4**, **5**, **2** and **7** (GSD, LNG, DSG, NET and ANGM, respectively) show relative binding affinities to SHBG of 17, 13, 5, 2.5 and 0%, respectively. We define molecule **8** as

Table 3
Relative binding affinities [1] for SHBG and five selected parameters[a]

Molecule	Binding affinity for SHBG (%)	PCA					EIM	
		η	Q_{17}	$F_5^{(o)}$	$F_7^{(o)}$	$F_9^{(o)}$	ρ''	Δ
8 (DHT)	100	5.578	3.10×10^{-2}	3.62×10^{-2}	1.34×10^{-3}	4.08×10^{-3}	−0.3532	0.1281
6 (GSD)	17	4.935	2.16×10^{-1}	7.47×10^{-1}	1.14×10^{-2}	1.32×10^{-2}	−0.0846	0.2282
4 (LNG)	13	4.988	1.94×10^{-1}	7.50×10^{-1}	1.43×10^{-2}	1.62×10^{-2}	0.5790	0.3193
5 (DSG)	5	5.210	1.94×10^{-1}	8.58×10^{-1}	1.57×10^{-2}	1.65×10^{-2}	0.6774	0.4009
2 (NET)	2.5	4.985	1.90×10^{-1}	7.48×10^{-1}	1.14×10^{-2}	1.35×10^{-2}	0.5786	0.3182
7 (ANGM)	0	4.664	1.98×10^{-1}	6.40×10^{-1}	1.20×10^{-2}	1.26×10^{-2}	0.2606	1.3958
1 (P)	–	5.002	-1.54×10^{-1}	7.51×10^{-1}	1.40×10^{-2}	1.05×10^{-2}	0.5278	0.2660
3 (NGM)	–	4.652	2.24×10^{-1}	6.43×10^{-1}	0.77×10^{-2}	0.94×10^{-2}	0.2997	1.3625

[a] Molecular hardness (η), net atomic charge in position 17 (Q_{17}) of SS and frontier orbital densities in positions 5 ($F_5^{(o)}$), 7 ($F_7^{(o)}$) and 9 ($F_9^{(o)}$) of SS for the progestogen contraceptives and related compounds. Binding affinity is not known for **1** and **3**. The last two columns list the two descriptors, ρ'' and Δ, defined by Eqs. (1) and (2) that are used for the EIM analysis.

Table 4
Relative oral progestational activities for 19 substituted 17α-acetoxyprogesterones [6]

Molecule	Oral progestational activities relative to norethisterone
1′ (17α-acetoxyprogesterone)	0.07
2′ (21-chloro-1,6-bisdehydro-17α-acetoxyprogesterone)	0.2
3′ (6α-nitro-17α-acetoxyprogesterone)	0.21–0.28
4′ (6β-chloro-17α-acetoxyprogesterone)	0.5
5′ (6α-fluoro-17α-acetoxyprogesterone)	1
6′ (21-fluoro-1,6-bisdehydro-17α-acetoxyprogesterone)	1
7′ (6α-bromo-17α-acetoxyprogesterone)	1
8′ (6α-methyl-17α-acetoxyprogesterone)	2–3
9′ (6α-chloro-17α-acetoxyprogesterone)	2–3
10′ (6α-bromo-1-hydro-17α-acetoxyprogesterone)	6
11′ (6α-fluoro-1-hydro-17α-acetoxyprogesterone)	6
12′ (1,6-bisdehydro-6α-fluoro-17α-acetoxyprogesterone)	8
13′ (1-hydro-6α-methyl-17α-acetoxyprogesterone)	8
14′ (6α-chloro-1-hydro-17α-acetoxyprogesterone)	8
15′ (6-methyl-1,6-bisdehydro-17α-acetoxyprogesterone)	10
16′ (6-methyl-6-hydro-17α-acetoxyprogesterone)	12
17′ (6α-fluoro-6-hydro-17α-acetoxyprogesterone)	15
18′ (6-chloro-1,6-bisdehydro-17α-acetoxyprogesterone)	35
19′ (6-chloro-6-hydro-17α-acetoxyprogesterone)	50

belonging to the high activity category and molecules **6**, **4**, **5**, **2** and **7** to the low activity category. The physicochemical parameters that are related with the binding affinity to SHBG are selected.

Since the number of steroids used in our study is quite small, we also take a set of 19 substituted 17α-acetoxyprogesterones (17α-AP) (Table 4), for which the relative oral progestational activities (OPA) [6] are known, to check the reliability of our results.

2. Methods

Molecular geometry was calculated with the semi-empirical Austin Model 1 (AM1) method [7]. Biological activity of a drug is believed to depend mainly on three different molecular properties: electronic, steric, and hydrophobic. Our strategy is to calculate as many physicochemical descriptors (parameters) as possible, using available software, since we do not know, ahead of time, which properties are more closely related to the biological activity that we are studying. We believe that some of the calculated descriptors correlate to the biological activity and some do not. Statistical methods are used to choose a set of the descriptors that correlate best to the biological activity. The calculated physicochemical parameters in the present work were: ionization potential (I), which was approximated by taking the negative value of the highest occupied molecular orbital (HOMO) energy (Koopmans' theorem); electron affinity (A), which was approximated by taking the negative value of the LUMO energy (Koopmans' theorem); molecular hardness ($\eta = (I - A)/2$) [8]; Mulliken eletronegativity ($\chi = (I + A)/2$); net atomic charge in the nth atom (Q_n); frontier indices [9] such as frontier electron density ($F_n^{(e)}$), frontier orbital density ($F_n^{(o)}$) and frontier radical density ($F_n^{(r)}$); van der Waals volume (VW) and area (VA) and molecular octanol–water partition coefficient ($\log P$). The quantum chemical parameters were obtained from the outputs of AM1 semi-empirical calculations. Some of them were calculated using the resultant molecular orbitals and their energies. The van der Waals volume and area were calculated using the SURF program [10]. The octanol–water partition coefficients ($\log P$) were calculated using parameters of substituent hydrophobicity [11]. Thus, a total of 45 molecular properties (descriptors) were calculated for each molecule studied.

The physicochemical parameters were correlated with the biological activity through the use of four different methods: (1) principal component analysis (PCA), (2) hierarchical clustering analysis (HCA), (3) neural networks (NN), (4) electronic indices method (EIM). It is desirable to apply different methods to the same problem in SAR. If all the different methods give identical or similar answers to the problem, it would be an indication that the answer obtained can be reliable.

PCA [12] is a useful exploratory tool, which maps samples through scores and individual variables by the loadings in a new vector space defined by the principal components. Score plots allow sample identification, checking if they are similar or dissimilar, typical or outliers. Also, it provides information about their groupings. From loading plots, the important variables can be identified and also the correlation pattern among them can be deduced. The first PC is generated in such a way that it has maximum correlation with all of the variables and usually accounts for a large portion of the total variance of the data. After removal of the first PC, a second PC is extracted which is completely uncorrelated (orthogonal) to the first one and accounts for the maximum possible remaining variance of the dataset. The procedure is repeated until all, or nearly all, of the original data has been utilized [13].

HCA [14], also an exploratory tool, is used to confirm the groupings previously identified by PCA. The primary goal of HCA is to emphasize the natural grouping of similar samples based on their closeness in the multidimensional space spanned by the variables. The results, qualitative in nature, are presented in the form of a dendogram, allowing visualization of clusters and of the correlation among samples. In HCA, the Euclidean distances between the samples are calculated and transformed into a similarity matrix whose elements are similarity indexes ranging from zero to one; a smaller distance means a larger index [13]. We used the computer programs called PIROUETTE [15] for PCA and HCA.

NN have been found to be suitable for data processing in which the relationship between the cause and its effects cannot be exactly defined. Thus, its use in biology-related responses is suggested [16]. The computer program PSDD [17], purchased from QCPE, was used for the NN calculations. The NN structure consists of three layers [16]. The value of a neuron in each layer is expressed by a sigmoid function. The back propagation method is used in PSDD. The process of calculation was carried out in a supervised manner.

The concept of the electronic densities of states (DOS — number of electronic states per energy unit) and local densities of states (LDOS–DOS calculated over a specific region or atom) has been used in studying properties of solid state matter. The notion of valence band and empty band, as well as the gap between them, provide useful information about their properties. In studying chemical reactivity and/or biological activity of molecules, one can expect the concepts like DOS (and LDOS) of occupied and unoccupied energy levels could also play some important role. Barone et al. [18] showed that the concept of DOS and LDOS could be successfully used to identify whether a specific polycyclic aromatic hydrocarbon (PAH) molecule would present (or not) carcinogenic activity (biological activity). The energy separation between the HOMO and the second HOMO was shown to relate closely to carcinogenic activity. The energy separation is inversely proportional to DOS. This method is called EIM. EIM has been applied to an extended class of molecules and compared with other methods [19,20]. The EIM method is based on one or two major descriptors, ρ and/or $\varDelta$. ρ is defined by Eq. (1)

$$\rho = 2\sum_{m=n_\mathrm{i}}^{n_\mathrm{f}} (|c_{m\mathrm{Level\,A}}|^2 - |c_{m\mathrm{Level\,B}}|^2) \quad (1)$$

where $c_{m\mathrm{Level}\,i}$ is mth atomic orbital coefficient in ith level molecular orbital (MO), the summation is carried over between initial (n_i) and final (n_f) atomic orbitals selected. Factor 2 corresponds to the occupation number of the concerned molecular orbitals. The ρ is a difference of electron density between two selected molecular orbitals, Level A and Level B, in a specified region, limited by n_i and n_f, of the molecule concerned. Level A is usually chosen as HOMO, while Level B is chosen as HOMO-1, one level below the HOMO. The second major EIM descriptor is $\varDelta$ which is defined by Eq. (2)

$$\varDelta = \varepsilon_{\mathrm{Level\,A}} - \varepsilon_{\mathrm{Level\,B}} \quad (2)$$

where $\varepsilon_{\mathrm{Level\,A}}$ is the molecular orbital energy of Level A (HOMO), while $\varepsilon_{\mathrm{Level\,B}}$ is that of Level B (HOMO-1). The calculations of the descriptors were carried out using Chem2Pac software [21], version 2.0.[2]

3. Results and discussion

3.1. Oral contraceptive activity

The process of the selection of the optimum descriptors in PCA is as follows. There are as many as 45 descriptors for each molecule. First, we visually analyze *biplot* diagrams projected on the monitor screen in order to reduce the number of descriptors to approximately one dozen. Then we apply PCA to the dozen descriptors selected. Usually no satisfactory discrimination of the molecules between high activity and low activity is obtained at this stage. Then, we discard one variable and apply PCA with the remaining variables. If the quality of the discrimination improves, we continue discarding the descriptors one by one till we get the optimum classification. The best separation was attained using only five descriptors (Table 1): the atomic charges in positions 10 (Q_{10}), 13 (Q_{13}) and 17 (Q_{17}) of the steroid skeleton (SS), the ionization potential (I) and the hardness (η). The PC scores graph is illustrated in Fig. 2. The high activity group is located on the left-hand side in the figure, while the low activity molecule, **1** (P), is on the right-hand side. These groupings are distinctively separated. The two principal components (PC1 and PC2) are given in Eqs. (3) and (4)

$$\mathrm{PC1} = 0.115I + 0.049\eta + 0.617Q_{10} + 0.481Q_{13} - 0.610Q_{17} \quad (3)$$

$$\mathrm{PC2} = 0.590I + 0.616\eta + 0.052Q_{10} - 0.490Q_{13} - 0.172Q_{17} \quad (4)$$

PC1 explains 50.1% of the variance and PC2 explains another 33.6%. Eq. (3) indicates that the three nuclear charges, Q_{10}, Q_{17} and Q_{13} are the major descriptors of PC1 while the outstanding descriptors of PC2 are η and I (Eq. (4)). Fig. 3 shows the loading graph for the five parameters. The three nuclear charges are mainly responsible for the separation of the six compounds seen in Fig. 2.

Fig. 4 shows the hierarchical clustering diagram for the set of molecules. It shows that molecule **1** has similarity zero to all other molecules of the group. Thus, the two different statistical methods, PCA and HCA, can classify the set of six molecules into the two categories: high activity and low activity, employing the five selected parameters.

The number of molecules studied (Table 1) is small, which may lead to questioning the usefulness of the results obtained. In order to investigate the reliability of our results, an external dataset of 19 substituted 17α-AP (Table 4), whose relative OPA (Clauberg assay) [6] is known, were added to the original OCA data. Their molecular structures are very similar to those in Fig. 1 and all the compounds in Table 4 have the acetoxy group (CH_3COO–) at position 17α. The six molecules (Table 1) and the nineteen molecules (Table 4) all belong to the family of “progestins”. Inhibition of ovulation was found to parallel closely the Clauberg

[2] http://www.ifi.unicamp.br/gsonm/chem2pac/.

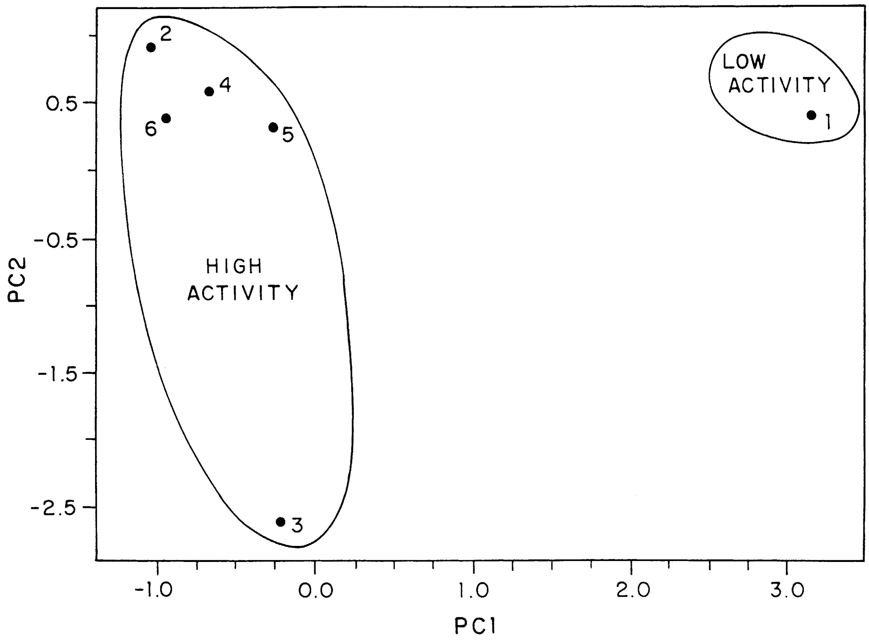

Fig. 2. Score graph of the first two principal components (PC1 and PC2) for Table 1 progestogens (oral contraceptive activity), using the five selected physicochemical parameters.

assays [22]. The structure–activity relationships of the substituted 17α-AP were previously investigated with methods similar to the ones described in this paper and the results have been published elsewhere [23].

We apply PCA to analyze the selected descriptors (I, η, Q_{10}, Q_{13} and Q_{17}) for 25 compounds (Tables 1 and 5). Fig. 5 shows the scores plot for the first two principal components for the expanded dataset where the 25 compounds are grouped into three regions, A, B, and C. Region A consists of compound **1** (low OCA), and mostly low OPA compounds, such as **1′**, **3′**, **4′**, **5′**. Region B consists mostly of high OPA compounds, such as **15′**–**17′** and **19′**. Region C consists of high OCA compounds, such as **2**–**6**. It is interesting to observe that the low OCA compound **1** and the low OPA compounds occupy the upper (above the dotted line) right-hand corner (region A) on the scores graph, while both the high

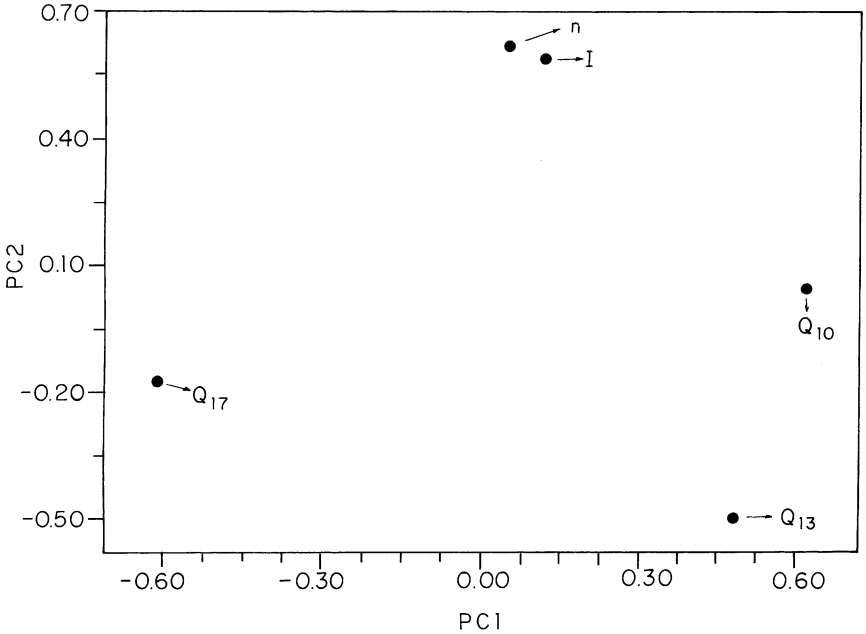

Fig. 3. PCA loadings of the five selected physicochemical parameters in progestogens (oral contraceptive activity).

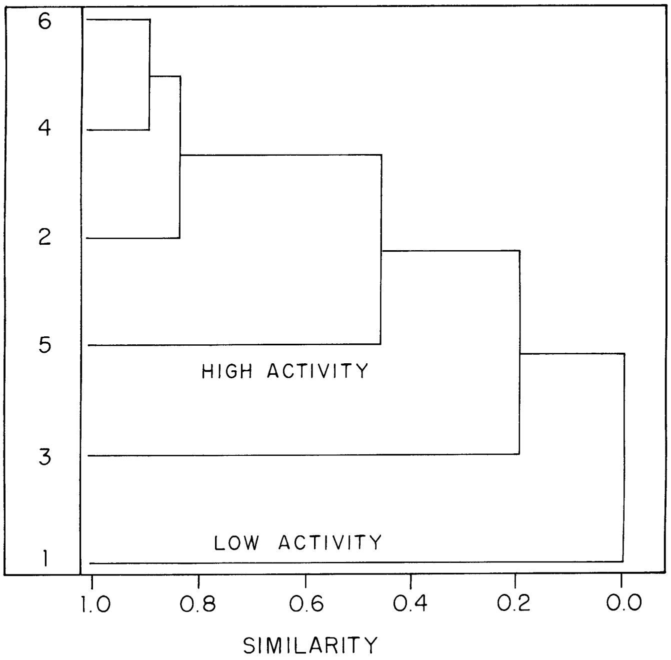

Fig. 4. Hierarchical clustering diagram for progestogens (oral contraceptive activity) using the five selected physicochemical parameters.

OPA compounds (region B) and high OCA compounds (region C) are located in the lower part (below the dotted line) of the figure. Regions B and C are divided distinctly along the PC1 axis. PC1 discriminates the 17α-AP molecules (regions A and B) from the non-acetoxyprogesterone molecules (region C). PC2 separates high activity molecules (regions B and C) from the low activity molecules (region A). The fact that the high OCA compounds (region C) as well as the high OPA compounds (region B) are located in the lower part of Fig. 5 might be an indication of the similarity between OCA and OPA. The two principal components (PC1 and PC2) are given in Eqs. (5) and (6)

$$\begin{aligned}\mathrm{PC1} &= 0.060I - 0.298\eta + 0.581Q_{10} \\ &\quad +0.561Q_{13} - 0.505Q_{17}\end{aligned} \tag{5}$$

Table 5
Five physicochemical descriptors[a]

Molecule	Q_{10}	Q_{13}	Q_{17}	I[b]	η[b]
1′	-1.83×10^{-2}	-3.32×10^{-2}	5.16×10^{-2}	10.051	4.990
2′	-5.10×10^{-3}	-2.67×10^{-2}	5.78×10^{-2}	9.441	4.381
3′	-1.20×10^{-2}	-3.07×10^{-2}	5.20×10^{-2}	10.567	4.937
4′	-1.39×10^{-2}	-3.14×10^{-2}	5.14×10^{-2}	10.370	4.958
5′	-1.08×10^{-2}	-2.09×10^{-2}	7.49×10^{-2}	10.229	4.934
6′	-6.90×10^{-3}	-2.40×10^{-2}	6.08×10^{-2}	9.453	4.407
7′	-1.37×10^{-2}	-3.06×10^{-2}	5.06×10^{-2}	10.290	4.996
8′	-1.94×10^{-2}	-3.29×10^{-2}	5.27×10^{-2}	10.025	4.989
9′	-1.59×10^{-2}	-2.73×10^{-2}	5.20×10^{-2}	10.245	4.948
10′	8.00×10^{-4}	-3.14×10^{-2}	5.13×10^{-2}	10.267	4.838
11′	1.30×10^{-3}	-3.22×10^{-2}	5.12×10^{-2}	10.252	4.834
12′	1.00×10^{-3}	-3.15×10^{-2}	5.08×10^{-2}	9.562	4.348
13′	-4.90×10^{-3}	-3.20×10^{-2}	4.96×10^{-2}	10.053	4.828
14′	5.00×10^{-4}	-3.19×10^{-2}	5.18×10^{-2}	10.253	4.828
15′	-6.20×10^{-3}	-3.25×10^{-2}	5.06×10^{-2}	9.356	4.371
16′	-2.03×10^{-2}	-3.22×10^{-2}	4.99×10^{-2}	9.275	4.392
17′	-1.38×10^{-2}	-3.18×10^{-2}	5.03×10^{-2}	9.479	4.372
18′	6.00×10^{-4}	-3.02×10^{-2}	4.83×10^{-2}	9.480	4.317
19′	-1.67×10^{-2}	-3.09×10^{-2}	4.97×10^{-2}	9.463	4.369

[a] Net atomic charges (Q_n), ionization potential (I) and hardness (η) for 19 substituted 17α-acetoxyprogesterones.
[b] In eV.

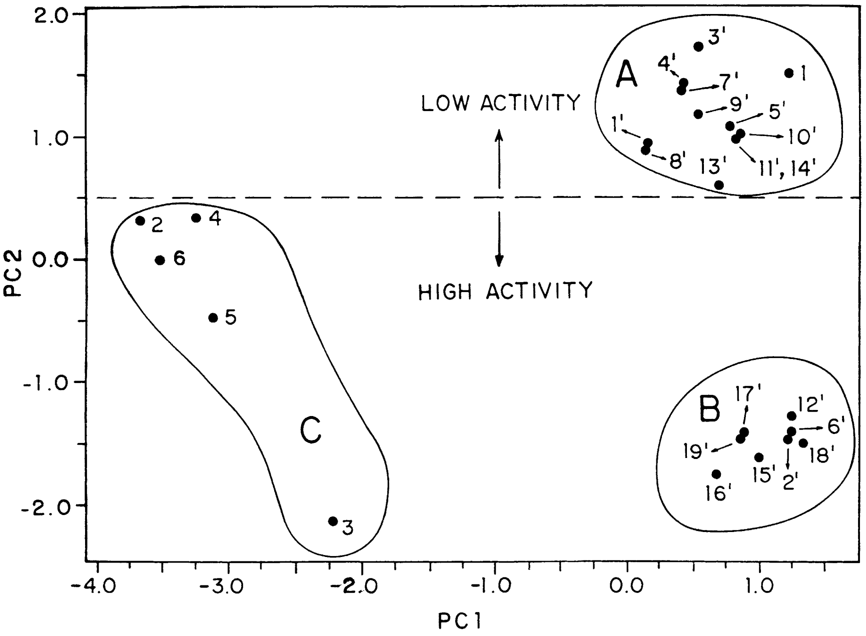

Fig. 5. Score graph of the first two principal components for the set composed of six progestogens (oral contraceptive activity, Table 1) and 19 17α-acetoxyprogesterones (Table 5) using the five selected physicochemical parameters.

$$\begin{aligned} \mathrm{PC2} &= 0.739I + 0.636\eta + 0.072Q_{10} \\ &\quad -0.004Q_{13} - 0.209Q_{17} \end{aligned} \qquad (6)$$

PC1 explains 55.1% of the variance and PC2 explains another 32.9%. Eq. (5) is similar, but not identical to Eq. (3), while Eq. (6) is similar, but not identical to Eq. (4). This is the reason why the two score graphs (Figs. 2 and 5), appear similar to each other as far as the relative distribution of the six compounds, **1**–**6** are concerned, but is not identical.

These results show that the five parameters, *I*, η, Q_{10}, Q_{13} and Q_{17} that were selected to classify the six molecules from Table 1 through PCA analysis are also useful to classify the expanded dataset. This is a demonstration that classification of samples with PCA can be meaningful even if the number of samples is as small as six. The five molecular properties chosen, out of the 45 descriptors, are efficient and effective descriptors for the classification of the type of molecules considered.

Since the five parameters, *I*, η, Q_{10}, Q_{13} and Q_{17}, seem to be useful in classifying progestins into high and low OCA groups by PCA, it is of interest to predict the OCA of molecules **7** and **8** (Fig. 1) whose OCA values are not available in the literature. The two molecules, **7** and **8**, were included with the six molecules in Table 1 to form a set of eight molecules. The PCA analysis was performed for the set of the eight molecules using the five parameters. The results of the PC analysis predicted that molecule **7** would have high activity and that molecule **8** would show low activity. These results are consistent with expectations. Molecule **7** is a metabolite of NGM (**3**, one of the three new progestogens) and it is expected to possess high OCA. Molecule **8** is not a progestin, but an androgen, which does not have progestational activity. Thus, molecule **8** should not have OCA activity.

In order to further investigate the soundness of the five descriptors selected by PCA, they were employed in NN analysis. The NN structure used for the calculations of OCA is listed in Table 6. The first layer consists of six neurons. The number of neurons of the first layer is equal to the number of molecular descriptors plus one, which is a bias. The same five descriptors selected by PCA (Table 1) were used in the NN calculations. The number of the second layer is five, which is a default value in the PSDD program. There are two neurons in the third layer to form two different patterns to represent the two categories, low activity and high activity. Table 7 lists the results of two types of NN calculations, RECALL and LONE (leave-one-out*)*. In RECALL, NN is asked to *recall* the data used for training NN with the training pattern. In LONE, NN is asked to predict the pattern of each member of the dataset, which are left out, one after another, during the training NN. The percent of correct classification was 100% both for RECALL and LONE.

Table 6
The NN structure used for NN calculations of oral contraceptive activity[a]

Layer	Neurons	α	θ
1	6		
2	5	1.0	0.0
3	2	1.0	0.0

[a] α is the non-linear parameter of the sigmoid functions and θ is a threshold value for a neuron (see Eq. (1) of [16]).

Table 7
NN classification of the molecules for oral contraceptive activity using the five descriptors listed in Table 1[a]

Molecule	Category	Training pattern	RECALL pattern	LONE pattern
1 (P)	1	1, 0	0.978, 0.022	Not applicable
2 (NET)	2	0, 1	0.009, 0.991	0.011, 0.989
3 (NGM)	2	0, 1	0.014, 0.986	0.042, 0.958
4 (LNG)	2	0, 1	0.013, 0.987	0.018, 0.982
5 (DSG)	2	0, 1	0.014, 0.987	0.027, 0.973
6 (GSD)	2	0, 1	0.009, 0.991	0.011, 0.989
Percent of correct classification			100	100

[a] Category 1 represents low activity, category 2 represents high activity. The RECALL pattern is the one calculated with the same input data as those used for the initial training of NN, LONE pattern refers to the results of the leave-one-out experiment. See Table 6 for the parameters used.

In Table 8, the results of PREDICTION of the category of OCA by NN are listed. NN was trained using one low activity compound, **1**, and two high activity compounds, **2** and **3**. The NN was used to predict the category of the remaining five compounds, **4**–**8**. The category of compounds **4**–**7** was predicted as category 2 (high activity), while that of the compound **8** as category 1 (low activity). This prediction corresponds to the classification previously made with PCA. NN calculations similar to those of Table 8 were carried out with nine different training sets. The results reveal that predictions were almost always correct. The category of **8** was predicted to be 2 instead of 1 twice. The prediction of the other compounds was always correct. This indicates that the model is robust. The five descriptors selected by PCA also work well in NN. Therefore, if other types of descriptors are arbitrary chosen, they do not work appropriately in NN predictions by LONE.

These results and those from Table 1 lead us to conclude that the OCA of the steroids is highly correlated with the three nuclear charges, Q_{10}, Q_{17} and Q_{13}, and to a lesser extent to η and I. Positive values of Q_{17} are seen in all the high OCA molecules (**2**–**6**) while Q_{17} is negative in the low OCA molecule **1**. All the high OCA molecules have ethynyl (–C≡CH) at the 17α position, whereas the low OCA molecule **1** has a hydrogen atom at 17α. The presence of (–C≡CH) at 17α seems to cause a positive nuclear charge on the carbon atom at position 17 of SS. The nuclear charge at position 10, Q_{10}, is negative for all the molecules, being approximately −0.02 for molecule **1** and in the range of −0.06 to −0.08 for the others. The presence of the methyl group at position 10 in molecule **1** (absent in high OCA molecules) greatly affects the magnitude of charge at this position. Values of Q_{13} are negative for the whole set of molecules. The high activity molecules have more negative Q_{13} values, in comparison to the low activity one. Low ionization potential is related to high contraceptive activity. In short, our results suggest that high contraceptive activity progesterones have (1) (large) positive nuclear charge at the position 17, (2) low negative nuclear charge at the positions 10 and 13, and (3) low molecular ionization potential.

Table 8
NN prediction of the category of oral contraceptive activity for the five molecules, **4**–**8**, using NN trained initially employing three molecules (**1**–**3**)

Molecule	Category	Training pattern	RECALL pattern	Predicted pattern
1 (P)	1	1, 0	0.978, 0.022	
2 (NET)	2	0, 1	0.016, 0.984	
3 (NGM)	2	0, 1	0.016, 0.984	
4 (LNG)	2			0.022, 0.987
5 (DSG)	2			0.053, 0.948
6 (GSD)	2			0.015, 0.985
7 (ANGM)	2			0.011, 0.989
8 (DHT)	1			0.950, 0.051

It is hoped that the information obtained above will be helpful in modeling better contraceptive progesterones. According to the set of molecules considered in this study, better contraceptive progesterones would satisfy results (1–3). It is beyond the scope of the present SAR work to reveal the detailed mechanisms of biological action of the studied molecules using the selected set of descriptors. The biological effects of steroids depend on many factors and can be very complex. The complete mechanism of action of these steroids is still unknown. With these limitations in mind, we will speculate about the implications of the selected descriptors with respect to contraceptive activity of the compounds, *assuming* that their activity depends mainly on the way they interact with their receptors.

Recently, Williams and Sigler [24] reported the 1.8 Å crystal structure of a progesterone-bound ligand binding domain (LBD) of the human progesterone receptor (PR). It clearly shows that the 3-keto oxygen in the A-ring of the steroid establishes a hydrogen-bonding network with the PR LBD. The role of the methyl-ketone substituent at C17 is less clear. It had been predicted previously that a progestin receptor site established an intimate specific contact with the A-ring, but a far less specific contact with the reminder of the steroid [25]. The result of the analysis of the crystal structure of Williams and Sigler coincides completely with the previous prediction. It is the A-ring of the steroid that plays an important role in the interaction between progestins and PR LBD. Three of the five descriptors, Q_{10}, I and η that were selected in the present SAR study (see Table 1) belong to the A-ring. The ionization potential, I, is due to the ionization event at HOMO which has a predominantly non-bonding character at the 3-keto oxygen [26]. The hardness, η, is related to HOMO and LUMO, which is an antibonding π-orbital of the carbonyl at position 3 (A-ring). Generally speaking, the reactivity of the electrons in HOMO is greater than any other electrons of the molecule against approaching electrophiles, such as a proton. Since the non-bonding orbital of the 3-keto oxygen in the A-ring is HOMO and it is localized, its electrons have the greatest capability to establish hydrogen bonding of any in the whole molecule. The ionization potential, I, selected as one of the descriptors is just HOMO energy with an opposite sign. The hardness, η, is also intimately related with the ability of the 3-keto oxygen to establish hydrogen bonding. As regards to the three charges selected, Q_{10}, Q_{13} and Q_{17}, the amount and sign of each charge may play an important role to establish optimum interaction between the steroid and its receptor. It is conceivable that there are opposing charges located counter to the positions 10, 13 and 17 of the steroid skeleton in the receptor.

Coulomb interaction between ligand and receptor, in addition to the hydrogen bonding, would contribute to stabilization of the ligand at the receptor site.

In the last two columns of Table 1, we present the values of ρ and Δ, defined by Eqs. (1) and (2), for use of the EIM analysis. After exploratory searching (analysing different molecular regions), we identified the region defined by the atomic orbitals of atom 10 (see Fig. 1) as the molecular region related to OCA. As can be seen in Table 1, progesterone, **1**, a low activity molecule, has $\rho < 0$, while most of the high activity molecules have $\rho > 0$. An exception occurs with the molecule **6** (GSD), that belongs to the group of high activity molecules but possess $\rho < 0$. Molecules **7** and **8** possess $\rho > 0$ and $\rho < 0$, respectively. According to the previous results, molecule **7** is related to high activity and molecule **8** to low activity. These results agree with the predictions obtained using PCA and NN. The second descriptor, Δ, of EIM was not useful for classification of the compounds in Table 1 for OCA (see the last column of Table 1). One of the five descriptors selected by PCA is charge at atom 10 (Q_{10}). The EIM method also selected the atom 10 as the key atom for the classifications. It seems that atom 10 plays an important role in determining OCA. Braga et al. [20] applied the EIM method to study relative OPA of 21 17α-AP's in which 19 molecules are identical to those in Table 4. They found the key region that correlates with relative OPA as that comprised by the four atoms at the positions 1, 2, 3, and 4 in SS (Fig. 1). This key region is different from that of the six OCA compounds in Table 1. This may be a consequence of the fact that OCA is not biologically identical to OPA, although they may be related closely to each other. The descriptor used in EIM is entirely different from the five descriptors selected by PCA. This demonstrates that the choice of descriptors in SAR studies depends on the type and method one employs.

3.2. Androgenic effect

Three parameters that best separate the highly active androgenic molecule, **4**; from the low androgenic activity molecules **1**–**3**, **5** and **6** were selected. They are listed in Table 2. They are the frontier radical density in position 7 ($F_7^{(r)}$) of the steroid skeleton, the frontier electron density in position 9 ($F_9^{(e)}$) and the frontier radical density in position 9 ($F_9^{(r)}$). The two principal components (PC1 and PC2) are described by Eqs. (7) and (8)

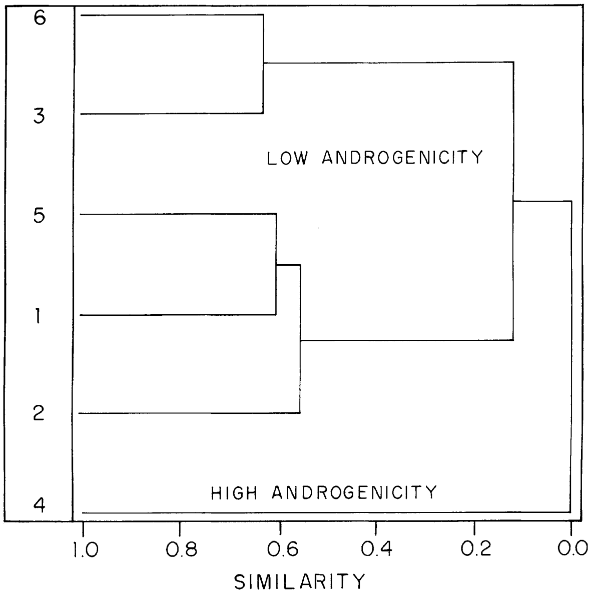

Fig. 6. Hierarchical clustering diagram for progestogen androgenic effects, using three physicochemical parameters from Table 2.

$$PC1 = 0.582F_9^{(e)} + 0.582F_7^{(r)} + 0.568F_9^{(r)} \quad (7)$$

$$PC2 = -0.420F_9^{(e)} - 0.383F_7^{(r)} + 0.823F_9^{(r)} \quad (8)$$

PC1 explains 91.5% of the variance and PC2 explains 5.6%. Eq. (7) indicates that the three frontier indices have about the same weight in PC1. The major descriptor of PC2 is $F_9^{(r)}$. The three frontier indices are mutually correlated to a high extent. The hierarchical clustering diagram (Fig. 6) shows similarity zero between molecule **4** and the rest of the group. This indicates that the highly active androgenic molecule, **4**, is separated from the rest of the molecules, which show low androgenic activity. The low androgenic activity molecules are grouped roughly into two subgroups: one consisting of **1**, **2**, **5**, the other consisting of **3**, **6**. Among all calculated descriptors, the three which were selected, are the best ones to discriminate these compounds. We believe that if more experimental data were available, a better discrimination might be attained, allowing the inclusion of other descriptors.

Again, the number of molecules treated for androgenic effect (Table 2) is small. Unfortunately, we could not find any additional sets of molecules, that have androgenic or similar effects to check our results, as was done for the OCA case. However, we have already shown that meaningful results can be obtained, even if the number of molecules is as small as six, so that it is reasonable to expect the same to be true for the case of androgenic effect.

The androgenic effect of two molecules, **7** and **8**, is not known. Since the three parameters, $F_7^{(r)}$, $F_9^{(e)}$ and $F_9^{(r)}$ seem to be useful in classifying the molecules into groups of high and low androgenic effect by PCA, it is possible to predict the androgenic effect of the two molecules using these three parameters, in a manner similar to that described for OCA. The PCA results predict that the two molecules, **7** and **8**, should have low androgenic effects. We expect that molecule **7** should possess a low androgenic effect because it is a metabolite of NGM (**3**). NGM is one of the three new progestogens which show only a small percentage of side effects.

We applied NN to analyze the androgenic effect of the molecules using the same three descriptors as those selected by PCA. The results of NN classification of the molecules for androgenic effect were as good as those obtained for contraceptive activity in Section 3.1. The NN predicted that the two molecules, **7** and **8**, should have low androgenic effects. This is in agreement with the prediction by PCA. These NN results imply that the three descriptors selected by PCA are appropriate for SAR study of androgenic effect of the compounds.

High values of $F_7^{(r)}$, $F_9^{(e)}$ and $F_9^{(r)}$ are related to high androgenic effects (Table 2). The high activity compound, **4** possesses the highest values of the frontier radical, $F^{(r)}$, and electron, $F^{(e)}$, densities at positions 7 and 9 of SS, in comparison to the rest of the group. If a steroid were to react with its reactant (or a receptor site of a macromolecule with which the steroid interacts), the frontier electrons located at positions 9 and 7 would react with the corresponding reactant (or receptor site). The reactant (or the receptor site) has either radical and/or electrophilic character. A low androgenic effect is found for the compound with low values of $F_7^{(r)}$, $F_9^{(e)}$ and $F_9^{(r)}$. Low androgenic effects are desired for an oral contraceptive, so, if one wants to synthesize a steroid with such characteristics, one would have to design the molecule in such a way that it has low values of $F_7^{(r)}$, $F_9^{(e)}$ and $F_9^{(r)}$. It is worth noting that the set of parameters selected for androgenic effect are completely different from the set of parameters selected for OCA. Androgenic activity seems to be mainly related with the frontier densities at positions 9 and 7, which belong to the B-ring, suggesting that this ring plays an important role in the interaction between the steroid and its receptor. In the case of OCA, it is the A-ring that interacts with the receptor through a hydrogen-bond-network. This might be an indication that the place and the nature of the receptor sites are different in OCA and androgenic effect for these steroids.

After exploratory searching (analyzing different molecular regions) for EIM analysis, we identified the region defined by atoms 15–17 (see Fig. 1) as the molecular region related to androgenic effect. From the last column of Table 2, we can observe that molecule **4** possesses $\rho' > 0$ (high activity androgenic molecule), while the low androgenic activity molecules **2**, **1**, **3**, **5** have $\rho' < 0$. The molecule **6** (GSD) again constitutes an exception. Molecules **7** and **8** have $\rho' < 0$. According to the previous results, they should be related to low androgenic effects. These results are in accordance with ones obtained from PCA and NN. EIM identified the D-ring as the molecular region that is related to androgenic activity. The PCA selected descriptors belong to the B-ring. Thus, in this case, the two different methods selected descriptors that belong to different regions of the molecules.

3.3. Binding affinity for SHBG

The five parameters that separate the high activity molecule **8** from the rest of the molecules were selected for PCA and they are listed in Table 3. They are frontier orbital densities in the positions 5 ($F_5^{(o)}$), 7 ($F_7^{(o)}$) and 9 ($F_9^{(o)}$) of the SS; atomic charge in position 17 (Q_{17}) of the SS and molecular hardness (η). The PC scores graph is presented in Fig. 7. Molecule **8** (DHT) has the highest binding affinity for SHBG and is located on the left hand side of the figure. It is well separated from the low activity molecules, which are on the right-hand side. The two principal components (PC1 and PC2) are given in Eqs. (9) and (10)

$$\begin{aligned}\mathrm{PC1} = {} & 0.468Q_{17} - 0.375\eta + 0.467F_5^{(o)} \\ & + 0.461F_7^{(o)} + 0.458F_9^{(o)}\end{aligned} \tag{9}$$

$$\begin{aligned}\mathrm{PC2} = {} & -0.159Q_{17} + 0.850\eta + 0.210F_5^{(o)} \\ & + 0.289F_7^{(o)} + 0.353F_9^{(o)}\end{aligned} \tag{10}$$

PC1 explains 88.3% of the variance and PC2 explains 10.4%. Eq. (9) indicates that the five selected descriptors have about the same weight in PC1.

The hierarchical clustering diagram is shown in Fig. 8. It shows similarity 0 of molecule **8** with respect to the rest of the molecules. The similarities among the low activity molecules are greater than 0.8, which is high, indicating that the selected parameters are well suited for the classification.

The binding affinity for SHBG of two molecules, **1** and **3**, is not known. Since the five parameters, Q_{17} η, $F_5^{(o)}$, $F_7^{(o)}$, $F_9^{(o)}$, seem to be useful in classifying the molecule into groups of high and low binding affinity by PCA, it is possible to predict the binding affinity for SHBG of these two molecules using the five parameters in a similar way as described in the previous sections. The results predicted by PC analysis indicated that both molecules, **1** and **3**, possess low binding affinities. In Table 3, we can observe that molecule **8**, which is an androgen, has a high binding affinity for SHBG while all the progestins show low binding affinity. Since **1** and **3** are progestins, we expect them to show low binding affinity.

We applied NN to analyze the binding affinity for SHBG of the molecules using the same five descriptors as those selected by PCA. The results of NN classification of the molecules for the binding affinities for SHBG were as good as those obtained for contraceptive activity. The NN predicted that two molecules, **1** and **3**, have low binding affinity for SHBG. This is in agreement with the prediction made by PCA. These NN results imply that the five descriptors

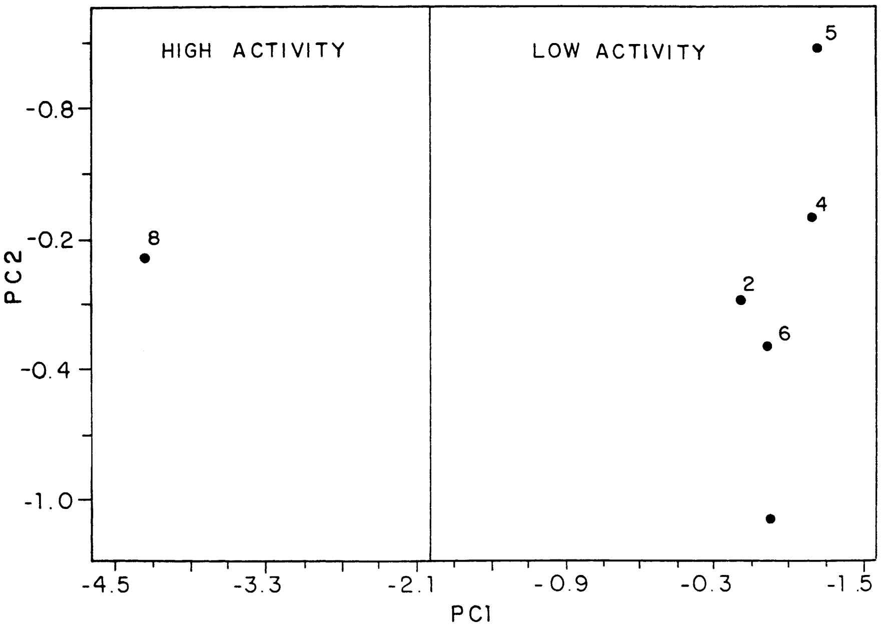

Fig. 7. Score graph of the first two principal components for progestogens for binding affinity for SHBG (Table 3) using five physicochemical parameters.

selected by PCA are appropriate for SAR study of the binding affinity for SHBG of the compounds.

Molecule **8** (DHT) which possesses the highest binding affinity to SHBG, has the highest value of η and the lowest Q_{17} of all (10 times lower than the others) (Table 3). In other words, a low value of η and high atomic charge at position 17 (Q_{17}) of SS are related to lower binding affinity to SHBG.

Frontier orbital densities at positions 5, 7 and 9 ($F_5^{(o)}$, $F_7^{(o)}$ and $F_9^{(o)}$) are also related to the binding affinity of SHBG. The lower the value of the frontier orbital densities in positions 5, 7 and 9, the higher the binding affinity to SHBG. Molecule **8** (DHT) possesses the highest binding affinity to SHBG while possessing the lowest values of $F_5^{(o)}$, $F_7^{(o)}$ and $F_9^{(o)}$. Molecules **6**, **4**, **5**, **2**, **7** that possess low binding affinity to SHBG, have values of $F_5^{(o)}$, $F_7^{(o)}$ and $F_9^{(o)}$ an order of magnitude (10 times) greater than molecule **8**, which has a high binding affinity. The frontier orbital density at a given atom ($F_n^{(o)}$) in a molecule is defined as the atomic orbital density in LUMO at the given atom. The frontier orbital density is a reactivity index with respect to nucleophiles [8]. When a steroid binds to its receptor, positions 5, 7 and 9 in the B-ring of SS may interact with the receptor sites that have nucleophilic character. A ketone at C3 is also important for optimum binding to SHBG [27]. The molecular hardness η, another selected parameter, is closely related to the keto group at C3. The hardness is approximated by $\eta \approx (\varepsilon_{\text{HOMO}} - \varepsilon_{\text{LUMO}})/2$ where $\varepsilon_{\text{HOMO}}$ and $\varepsilon_{\text{LUMO}}$ are the orbital energies of HOMO and LUMO, respectively. The HOMO and LUMO are localized in the keto at C3 [26]. A 17β-OH group is absolutely required for strong binding to SHBG [27]. The fact that Q_{17}, the atomic charge at 17, was selected in Eq. (5) as an important descriptor is related to this observation. Descriptor η is related to position 3 and Q_{17} is related to position 17. The correlation coefficient between η and Q_{17} is −0.83, which is fairly high. This may be a consequence of the long-range interactions between the 3 and 17 positions [28].

The last two columns of Table 3 summarize the EIM calculations for the eight steroids indicated in Fig. 1. After exploratory searching (analyzing different molecular regions), we identified the region defined by the atoms 5–10 (see

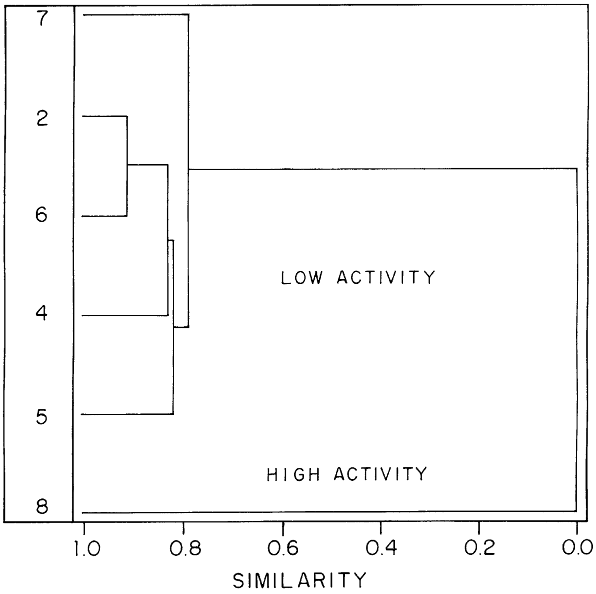

Fig. 8. Hierarchical clustering diagram for progestogen binding affinities for SHBG, using five physicochemical parameters from Table 3.

Fig. 1) as the molecular region related to binding affinity for SHBG. We can observe that molecule **8**, which shows the highest binding affinity, possesses $\rho'' < 0$, while the low binding affinity molecules **4**, **5**, **2**, **7** have $\rho'' > 0$. Molecule **6** (GSD) continues to constitute an exception. Molecules **1** and **3** have $\rho'' > 0$ and according to the previous results, they are related to low androgenic effects. These results are in agreement with the ones obtained from PCA and NN. The high activity molecule, **8**, has a smaller energy separation (Δ) than any other low activity molecule (Table 3). The energy separation (Δ) also serves as a useful discriminatory descriptor in case of binding affinity for SHBG of the compounds. The region selected by EIM belongs to the B-ring. Three descriptors out of the five selected by PCA belong also to the B-ring.

A comparison between Tables 1–3 reveals that the three sets of selected molecular parameters are different from each other. Except for Q_{17}, no parameters were selected more than once in the three sets. This might result from the fact that the nature of the receptors and the mode of interaction between the steroids and the receptors for the three different biological activities are different from each other. The mode of interaction between the receptor and the progestogen must depend on the type of biological activity. The receptor site and/or the mode of interaction between the receptor and the steroid for OCA is different from that for androgenic effect, which in turn is different from that for binding affinity to SHBG. An efficient oral contraceptive should have a low androgenic effect as well as low binding affinity to SHBG. Combining the results of the investigations on these three different biological activities (OCA, androgenic effect, and binding affinity to SHBG), we can have some insight into the nature of an efficient oral contraceptive progestogen.

4. Conclusions

The use of the PCA method, together with the parameters calculated with the semi-empirical AM1 method, enabled us to study the structure–activity relationship of contraceptive progestogens. The oral contraceptive activities of the progestogens (Fig. 1) are associated mainly with the atomic charges in positions 10 (Q_{10}), 13 (Q_{13}) and 17 (Q_{17}) of the steroid skeleton (SS), the ionization potential (I) and the hardness (η). The androgenic effect is associated with the frontier radical and electron densities at positions 7 and 9 ($F_7^{(r)}$, $F_9^{(r)}$ and $F_9^{(e)}$) of the steroid skeleton. The binding affinity for SHBG is mainly related with Q_{17}, with molecular hardness (η) and with the frontier orbital densities at positions 5, 7 and 9 ($F_5^{(o)}$, $F_7^{(o)}$ and $F_9^{(o)}$) of the steroid skeleton. Three different sets of descriptors were found to correlate with the three different biological activities, indicating that the interaction between the receptor and the progestogen and/or mode of action must depend on the type of biological activity.

Exactly the same descriptors as the ones selected by PCA were employed for SAR analysis of the contraceptive progestogens using two other methods: HCA and NN. Both HCA and NN correctly classified high activity molecules from low activity ones. Thus, those descriptors selected by PCA work well for the other two classification methods.

The EIM method is an entirely different approach to SAR analysis from the PCA, HCA and NN methods. The descriptors used in EIM are always ρ (Eq. (1)) and/or Δ (Eq. (2)). Using the sign of one or both of these, it was possible to discriminate high activity molecules from low activity molecules in the three different types of activities studied, with the exception of GSD. The type of descriptors that are useful in SAR analysis depend upon many factors, and thus, are usually not unique.

Acknowledgements

This work was financially supported by FAPESP, CAPES and CNPq. Computing facilities provided by CENAPAD/SP are acknowledged. Thanks are also due to Dr. Douglas S. Galvão for letting us use Chem2Pac software and for stimulating discussions related to the EIM method.

References

[1] K. Fotherby, A.D.S. Caldwell, New progestogens in oral contraception, Contraception 49 (1994) 1–32.

[2] N.J. Alexander, Future contraceptives, Scientific Am. 273 (1995) 136–141.

[3] C.M.H. Coenen, C.M.G. Thomas, G.F. Borm, J.M.G. Hollanders, R. Rolland, Changes in androgens during treatment with four low-dose contraceptives, Contraception 53 (1996) 171–176.

[4] R.W. Rebar, K. Zeserson, Characteristics of the new progestogens in combination oral contraceptives, Contraception 44 (1991) 1–11.

[5] J.F. Dunn, B.C. Nisula, D. Rodbard, Transport of steroid-hormones — binding of 21 endogenous steroids to both testosterone-binding globulin and corticosteroid-binding globulin in human plasma, J. Clin. Endocr. Metab. 53 (1981) 58–68.

[6] C.W. Shoppee, Chemistry of the Steroids, 2nd Edition, Butterworth, London, 1964, p. 185.

[7] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, The deve- lopment and use of quantum-mechanical molecular-models.76: AM1— a new general-purpose quantum-mechanical molecular-model, J. Am. Chem. Soc. 107 (1985) 3902–3909.

[8] R.G. Parr, R.G. Pearson, Absolute hardness-companion parameter to absolute electronegativity, J. Am. Chem. Soc. 105 (1983) 7512–7516.

[9] K. Fukui, Theory of orientation and stereo selection, in: K. Hafner, J.-M. Lehn, C.W. Rees, P.R. Schleyer, B.M. Trost, R. Zahradnik (Eds.), Reactivity and Structure Concepts in Organic Chemistry, Vol. 2, Springer, Berlin, 1975, p. 39.

[10] A.C. Gaudio, Y. Takahata, Calculation of molecular-surface area with numerical factors, Comput. Chem. 16 (1992) 277–284.

[11] G.G. Nys, R.F. Rekker, Concept of hydrophobic fragmental constants (*f*-values). 2. Extension of its applicability to calculation of lipophilicities of aromatic and heteroaromatic structures, Eur. J. Med. Chem. Clin. Ther. 9 (1974) 361–375.

[12] K. Fukunaga, W.L.G. Koontz, Application of Karhunen–Loeve expansion to feature selection and ordering, IEEE T. Comput. C-19 (1970) 311–318.

[13] K.R. Beebe, R.J. Pell, M.B. Seasholtz, Chemometrics: A Practical Guide, Wiley, New York, 1998.

[14] B.R. Kowalski, C.F. Bender, Pattern-recognition — powerful approach to interpreting chemical data, J. Am. Chem. Soc. 94 (1972) 5632–5639.

[15] Pirouette multivariate data analysis for IBM PC systems, version 2.7, 2000, Infometrix, Seattle, WA.

[16] T. Aoyama, Y. Suzuki, H. Ichikawa, Neural networks applied to structure–activity relationships, J. Med. Chem. 33 (1990) 905–908.

[17] H. Ichikawa, PSDD: perceptron-type neural network simulator, QCPE 615 (2001).

[18] P.M.V.B. Barone, A. Camilo Jr., D.S. Galvão, Theoretical approach to identify carcinogenic activity of polycyclic aromatic hydrocarbons, Phys. Rev. Lett. 77 (1996) 1186–1189.

[19] R. Vendrame, R.S. Braga, Y. Takahata, D.S. Galvão, Structure–activity relationship studies of carcinogenic activity of polycyclic aromatic hydrocarbons using calculated molecular descriptors with principal component analysis and neural network methods, J. Chem. Inform. Comput. Sci. 39 (1999) 1094–1104.

[20] R.S. Braga, R. Vendrame, D.S. Galvão, Structure–activity relationship studies of substituted 17-acetoxyprogesterone hormones, J. Chem. Inform. Comput. Sci. 40 (2000) 1377–1385.

[21] M. Cyrillo, D.S. Galvão, Chem2Pac: a computational chemistry integrator for windows, EPA News Lett. 67 (1999) 31–34.

[22] H.J. Ringold, E. Batres, A. Bowers, J. Edwards, J. Zderic, Steroids. 127. 6-halo progestational agents, J. Am. Chem. Soc. 81 (1959) 3485–3486.

[23] R. Vendrame, Y. Takahata, Structure–activity relationship (SAR) of substituted 17alpha-acetoxyprogesterones studied with principal component analysis and neural networks using calculated physicochemical parameters, J. Mol. Struct. (THEOCHEM) 489 (1999) 55–66.

[24] S.P. Williams, P.B. Sigler, Atomic structure of progesterone complexed with its receptor, Nature 393 (1998) 392–396.

[25] W.L. Duax, J.F. Griffin, C.M. Weeks, in: M. Sluyser (Ed.), Interaction of Steroid Hormone Receptors with DNA, Ellis Horwood, Chichester, UK, 1985, p. 83.

[26] Y. Takahata, R. Vendrame, Ionization energies and electron affinities of some steroids calculated with the semi-empirical HAM/3 method, J. Mol. Struct. (THEOCHEM) 391 (1997) 169–178.

[27] G.R. Cunningham, D.J. Tindall, T.J. Lobl, J.A. Campbell, A.R. Means, Steroid structural requirements for high-affinity binding to human sex steroid binding-protein (SBP), Steroids 38 (1981) 243–262.

[28] T. Cvitaš, B. Kovac, L. Paša-Tolic, B. Rušcic, L. Klasinc, J.V. Knop, N.S. Bhacca, S.P. McGlynn, Photoelectron-spectra, electronic-structure and long-range electronic interaction in some steroids, Pure Appl. Chem. 61 (1989) 2139–2150.