

## Note

### Synthesis and QSAR studies of 16-(3-methoxy-4-substituted benzylidene) androstene derivatives as anticancer agents

Sonal Dubey\*<sup>a</sup>, Parmeet Kaur<sup>a</sup>, Dharam Paul Jindal<sup>b</sup>,  
Yalamanchili Darji Satyanarayan<sup>a</sup> & Poonam Piplani<sup>b</sup>

<sup>a</sup>K.L.E. College of Pharmacy, II-Block Rajajinagar,  
Bangalore 560 010, India

<sup>b</sup>University Institute of Pharmaceutical Sciences, Panjab  
University, Chandigarh 160 014, India

E-mail: drsonaldubey@gmail.com

Received 11 March 2008; accepted (revised) 25 March 2010

In a systematic effort aimed at identifying new steroidal cytotoxic agents with potent antiproliferative activity against cancer cells and developing their QSAR models, a series of 16-(3-methoxy-4-substituted benzylidene)androst-5-ene derivatives have been synthesized. The selected compounds are evaluated for antineoplastic activity against a panel of three human cell lines—breast, CNS and lungs at NCI, Bethesda, USA. The results presented herein indicate that compound **15-18, 21, 22, 25-30** are active anticancer agents. The QSAR investigation with multiple linear regression analysis has been applied to find a correlation between different calculated physicochemical parameters of these compounds and biological activity. Application of datasets by using CODESSA software has led to QSAR equations based on the 3 descriptors. The significant QSAR models have been obtained with  $R^2$  values which range from 0.9692-0.8225 and good predictive performance ( $q^2$  range: 0.9264-0.7121). These models are expected to be useful for the anticancer screening of androstene derivatives having substitution at position 3,16 and 17 of steroid nucleus.

**Keywords:** Androstane, anticancer, *in vitro* activity, QSAR, steroids

Hormonal agents have a confirmed role in the management of cancer like breast and prostate<sup>1</sup>. Certain naturally occurring tumoricidal steroidal glycosides have been reported from *Tribulus terrestris*<sup>2</sup>, *Solanum incanum*<sup>3</sup>, *Momordica dioica*<sup>4</sup> (**1**), *etc.* Many steroidal derivatives have also been prepared synthetically and some are deemed as drugs *e.g.* formastane<sup>5,6</sup> (**2**) and exemestane<sup>7-9</sup> (**3**), *etc.* Certain steroidal nitrosoureas have been found to compete with estradiol for binding to cytosolic estrogen receptor in rat uterus<sup>10,11</sup>. Certain novel steroid-linked conjugates of 17 $\beta$ -[N-[N'-(2-chloroethyl)-N'-nitroso]carbonyl] aminoacids have been

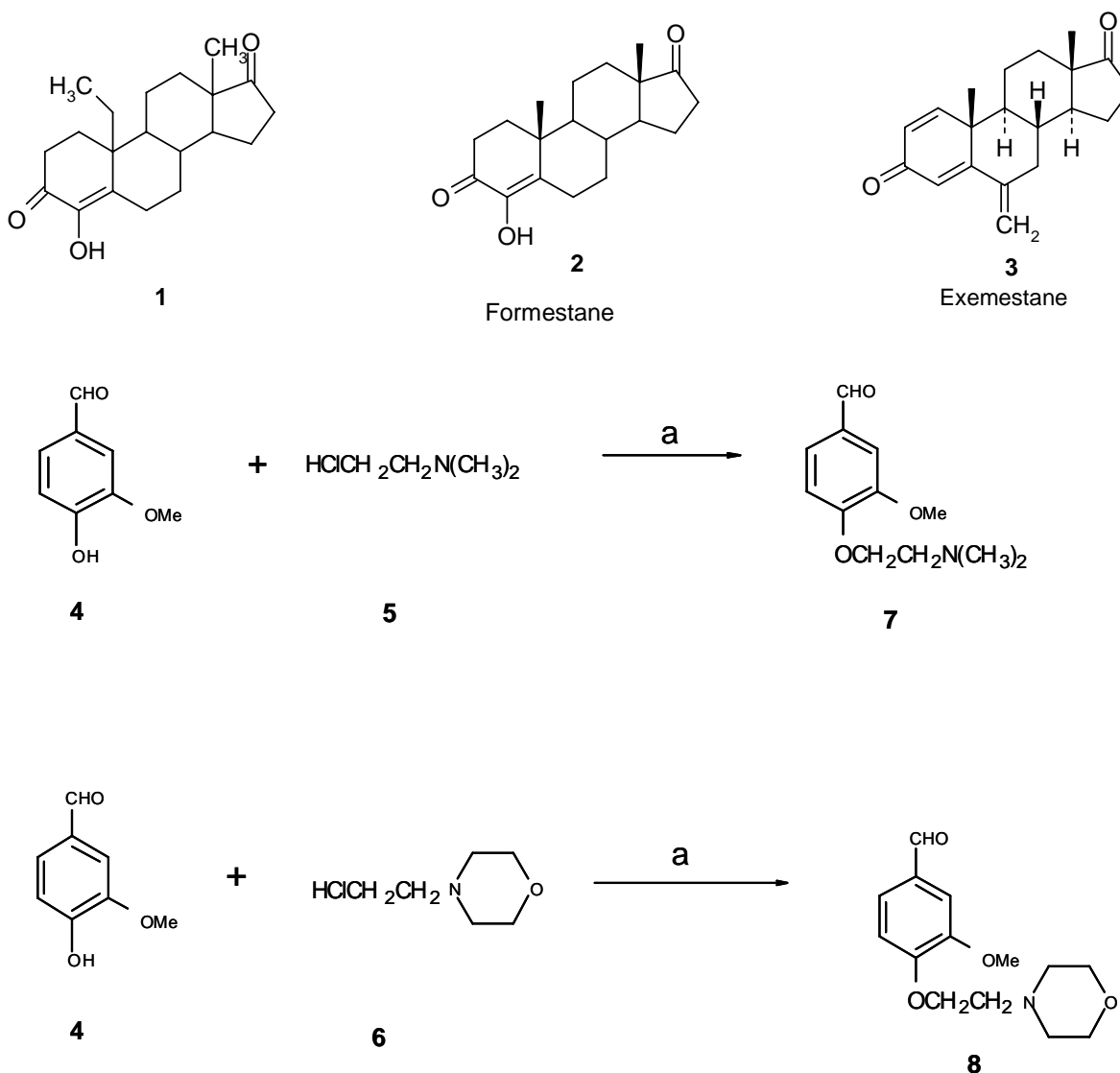
synthesized *e.g.* 17 $\beta$ -CNC-glycyl-19-nortestosterone, which showed antineoplastic activity by binding to progesterone receptor, while 17 $\beta$ -CNC-alanyl-19-nortestosterone, possesses affinity for androgen receptor<sup>12</sup>. Some of the steroidal arylidene oxime derivatives synthesized in the laboratories of University Institute of Pharmaceutical Sciences, Panjab University were found to be active against one or more human tumor cell lines, when tested for *in vivo* hollow fiber assay at NCI, USA<sup>13</sup>.

In the present study, certain novel 16-(3-methoxy-4-substituted benzylidene) androstenes have been designed and synthesized. The selected synthesized compounds were evaluated for their *in vitro* anticancer activity. The quantitative structure activity relationship (QSAR) investigations have been performed in order to have a better understanding of different physicochemical properties. Multiple linear regression analysis was used as the statistical tool to find the correlation between different physicochemical parameters of the compounds and their biological activity. The models were then validated for their predictive ability. Beside the explanatory ability of these QSAR models, they could be used to predict the potency of yet to be synthesized compounds and to suggest ideas for further synthesis of new molecules with enhanced activity.

## Results and Discussion

### Chemistry

A general outline for the synthesis of these 16-*para*-substituted benzylidene derivatives is given in **Schemes I** and **II**. First the vanillin was converted to 3-methoxy 4-substituted benzaldehyde (**Scheme I**). As per **Scheme II** these substituted aldehydes **7** and **8** were then subjected to aldol condensation<sup>14</sup> with dehydroepiandrosterone **9** to obtain the compounds **10** and **21** respectively. This was followed by Oppenauer oxidation<sup>15</sup> of **10** and **21** using cyclohexanone/toluene system which afforded 3,17-diones **11** and **22**. On treatment with pyrrolidine in methanol, the  $\alpha,\beta$ -unsaturated ketones **11** and **22** gave enamines **12** and **23** respectively. Both the enamines were reduced with sodium borohydride to give **13** and **24**, which were then subjected to esterification to get **14** and **25**.



**Scheme I** — (a) Ethylmethyl ketone/ potassium carbonate; reflux at 110°C.

Reduction of **10** and **21** with sodium borohydride yielded the compounds **19** and **30** and subsequent acetylation with acetic anhydride yielded compounds **20** and **31** respectively. The direct acetylation of **10** and **21** with acetic anhydride in dry pyridine afforded **17** and **28**. A detailed account of the synthesis of 17-oximino **15** and **26**, 3-acetoxy-17-oximino **18** and **29** and 3,17-dioximino derivatives **16** and **27** has already been reported<sup>16,17</sup>.

All the physical constants of the synthesized compounds are given in **Table I**.

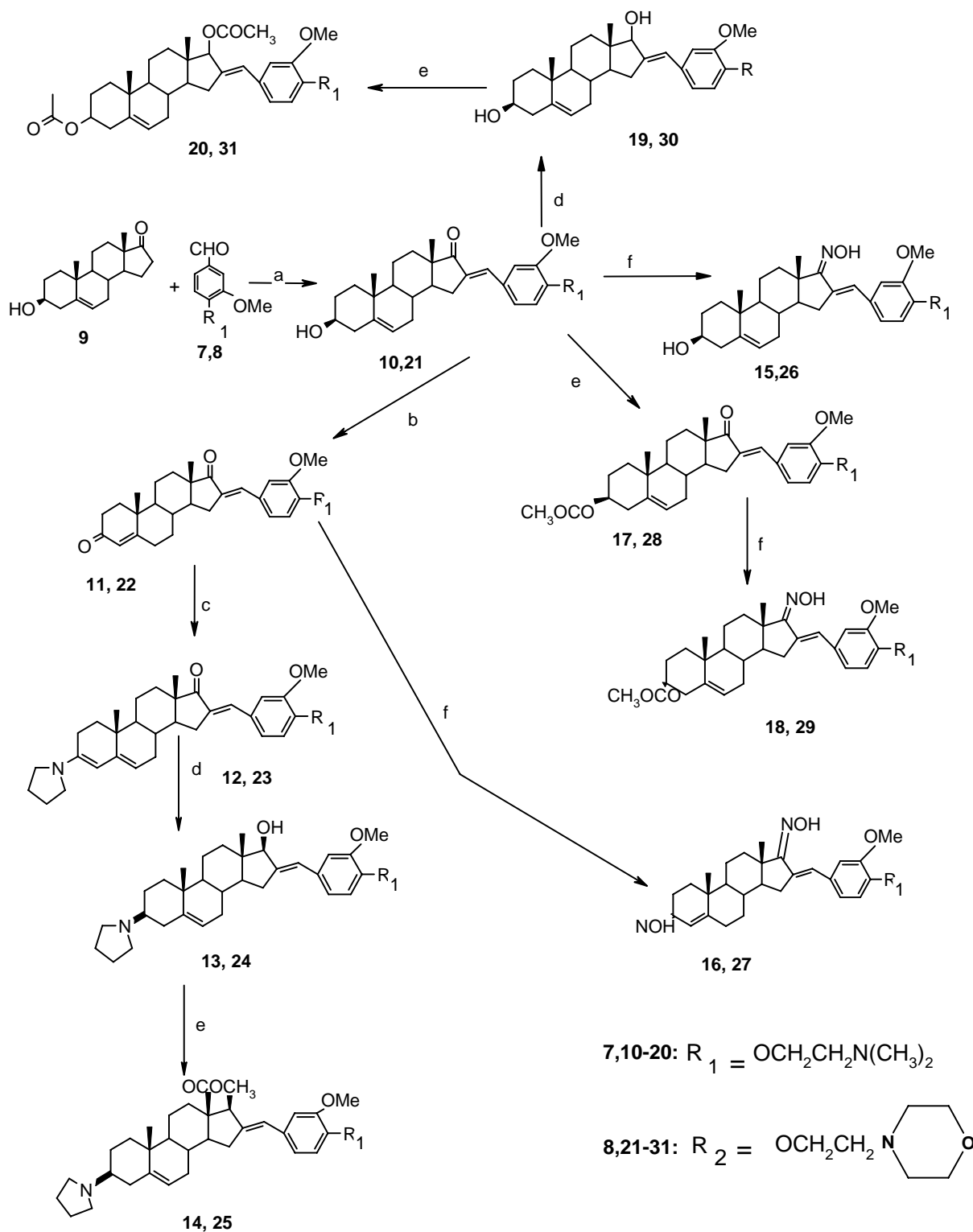
#### Anticancer activity

Out of the fifteen compounds – **10**, **11**, **13**, **15-18**, **21**, **22**, **25-30** tested for their antineoplastic activity

against three human cell lines (lung, breast and CNS), twelve – **15-18**, **21**, **22**, **25-30** were found to possess moderate to high activity (**Table II**).

#### QSAR studies

The above data set was split into TR, including 12 compounds, and TS, including 3 compounds **10**, **18** and **25**. On the basis of the TR, several QSAR equations were built, each containing different number of descriptors. The model selected as the best one included the ratio of compounds to descriptors 4:1 (Ref. 18); together with the largest value of  $q^2$  and  $R^2$ . Based on these the best QSAR models derived for each cell line are:



Reagents and Conditions: (a) Sodium hydroxide, shaken at RT; (b) Cyclohexanone, aluminium isopropoxide; (c) Pyrrolidine, methanol, reflux; (d) Sodium borohydride, methanol at RT; (e) Acetic anhydride, pyridine reflux; (f) hydroxylamine hydrochloride/ sodium acetate trihydrate, reflux in aldehyde free ethanol.

**Scheme II** — Outline of synthetic procedure for compounds **10-31**

**Table I** — Physical characterization data of the compounds

| Compd     | Yield (%) | m.p. (°C) | Mol. Formula (Mol. Wt.)  | Calcd (Found)    |               |              | <sup>1</sup> H NMR (δ, ppm)  | MS (m/z)              |
|-----------|-----------|-----------|--|------------------|---------------|--------------|--|-----------------------|
|           |           |           |  | C                | H             | N            |  |                       |
| <b>10</b> | 54.52     | 190-94    | C <sub>31</sub> H <sub>43</sub> NO <sub>4</sub><br>(493)               | 75.46<br>(75.01) | 8.72<br>8.54  | 2.84<br>2.97 | 2.43 [s, 6H, N-(CH <sub>3</sub> ) <sub>2</sub> ], 2.9 (t, 2H, -CH <sub>2</sub> -N<),<br>3.88 (s, 3H, -OCH <sub>3</sub> ), 4.19 (t, 2H, -OCH <sub>2</sub> -),<br>5.39 (d, 1H, 6-CH), 7.38 [s, 1H, 16-vinyl-H] | 493[M <sup>+</sup> ]  |
| <b>11</b> | 99        | 78-82     | C <sub>31</sub> H <sub>41</sub> NO <sub>4</sub><br>(491)               | 75.76<br>(74.72) | 8.35<br>7.56  | 2.85<br>2.17 | 5.76 (s, 1H, 4-CH), 7.40 {s, 1H, 16-vinyl-H}   | 491[M <sup>+</sup> ]  |
| <b>12</b> | 72.20     | 112-16    | C <sub>35</sub> H <sub>48</sub> N <sub>2</sub> O <sub>3</sub><br>(544) | 77.21<br>(77.00) | 8.82<br>8.56  | 5.15<br>4.98 | 3.15 (br, 4H, N-methylenes of pyrrolidine function),<br>5.40 (s, 2H, 4-CH and 6-CH protons).   | 544[M <sup>+</sup> ]  |
| <b>13</b> | 86.86     | 178-82    | C <sub>35</sub> H <sub>52</sub> N <sub>2</sub> O <sub>3</sub><br>(548) | 76.6<br>(75.99)  | 49.49<br>9.34 | 5.11<br>4.97 | 2.58 (br, 4H, -N-methylenes of pyrrolidine),<br>4.05 (s, 1H, 17α-H), 5.36 (d, 1H, 6-CH),<br>6.86-6.96 (m, 3H, aromatic H).   | 548 [M <sup>+</sup> ] |
| <b>14</b> | 51.08     | 122-26    | C <sub>37</sub> H <sub>54</sub> N <sub>2</sub> O <sub>4</sub><br>(590) | 75.21<br>(74.81) | 9.21<br>8.46  | 4.74<br>4.48 | 2.21 (s, 3H, 3-OCOCH <sub>3</sub> ), 5.35 (m, 2H, 6-CH and 17α-H).   | 590[M <sup>+</sup> ]  |
| <b>17</b> | 66.18     | 144-47    | C <sub>33</sub> H <sub>45</sub> NO <sub>5</sub><br>(535)               | 74.02<br>(73.85) | 8.41<br>7.64  | 2.42<br>2.26 | 2.05 (s, 3H, 3-OCOCH <sub>3</sub> ), 7.38 {s, 1H, 16-vinyl-H}  | 535 [M <sup>+</sup> ] |
| <b>19</b> | 44.82     | 140-45    | C <sub>31</sub> H <sub>45</sub> NO <sub>4</sub><br>(495)               | 75.15<br>(75.11) | 9.09<br>8.71  | 2.83<br>2.79 | 3.53 (m, 1H, 3α-H), 4.04 (s, 1H, 17α-H)  | 495[M <sup>+</sup> ]. |
| <b>20</b> | 51.62     | 90        | C <sub>35</sub> H <sub>49</sub> NO <sub>6</sub><br>(579)               | 72.54<br>(71.99) | 8.46<br>7.98  | 2.42<br>2.23 | 2.04(s, 3H, 3-OCOCH <sub>3</sub> ), 2.20(s,3H, 17-OCOCH <sub>3</sub> )   | 579[M <sup>+</sup> ]  |
| <b>21</b> | 63.20     | 174-76    | C <sub>33</sub> H <sub>45</sub> NO <sub>5</sub><br>(535)               | 74.02<br>(73.72) | 8.41<br>8.71  | 2.61<br>2.04 | 2.64 (br, 4H, N-methylenes of morpholino)<br>3.76 (m, 4H, O-methylenes of morpholino),<br>6.93-7.14 (aromatic H), 7.38 {s, 1H, 16-vinyl-H}   | 535 [M <sup>+</sup> ] |
| <b>22</b> | 85.34     | 62-66     | C <sub>33</sub> H <sub>43</sub> NO <sub>5</sub><br>(533)               | 74.30<br>(73.67) | 8.07<br>7.71  | 2.63<br>2.45 | 5.77 (s, 1H, 4-CH), 7.40 {s, 1H, 16-vinyl-H}   | 533[M <sup>+</sup> ]  |
| <b>23</b> | 77.31     | 105-08    | C <sub>39</sub> H <sub>50</sub> NO <sub>4</sub><br>(596)               | 76.72<br>(76.53) | 8.20<br>8.14  | 4.59<br>3.97 | 4.45 (br, 4H, N-methylenes of pyrrolidine),<br>5.10 (s, 1H, 4-CH), 5.25 (d, 1H, 6-CH),   | 610[M <sup>+</sup> ]  |
| <b>24</b> | 67.77     | 184-88    | C <sub>39</sub> H <sub>54</sub> N <sub>2</sub> O <sub>4</sub><br>(614) | 76.22<br>(76.05) | 8.79<br>8.25  | 4.56<br>4.23 | 4.05 (d, 1H, 17α-H), 5.37 (m, 1H, 6-CH),<br>6.45 {s, 1H, vinyl-H}  | 614 [M <sup>+</sup> ] |
| <b>25</b> | 46.68     | 130-35    | C <sub>41</sub> H <sub>56</sub> N <sub>2</sub> O <sub>5</sub><br>(656) | 75.00<br>(74.87) | 8.54<br>8.01  | 4.27<br>3.97 | 2.21 (s, 3H, 17-OCOCH <sub>3</sub> ),<br>2.59 (m, 8H, N-methylenes of morpholino and pyrrolidine),<br>6.85-6.93 (3H, aromatic protons)   | 656[M <sup>+</sup> ]  |
| <b>28</b> | 51.01     | 75-78     | C <sub>35</sub> H <sub>47</sub> NO <sub>6</sub><br>(577)               | 72.79<br>(72.12) | 8.15<br>8.01  | 2.43<br>2.36 | 2.04 (s, 3H, 3-OCOCH <sub>3</sub> ), 7.38 {s, 1H, vinyl-H}   | 577[M <sup>+</sup> ]  |
| <b>30</b> | 69.74     | 130       | C <sub>33</sub> H <sub>47</sub> NO <sub>5</sub><br>(537)               | 73.74<br>(73.72) | 8.75<br>7.70  | 5.61<br>4.99 | 5.35 (d, 1H, 6-CH), 6.86-6.95 (m, 3H, 2-CH, 5-CH and<br>6-CH aromatic protons).  | 537[M <sup>+</sup> ]  |
| <b>31</b> | 34.6      | 165-70    | C <sub>37</sub> H <sub>51</sub> NO <sub>7</sub><br>(621)               | 71.49<br>(71.00) | 8.21<br>7.86  | 2.25<br>2.04 | 2.04 (s, 3H, 3-OCOCH <sub>3</sub> ).   | 621[M <sup>+</sup> ]  |

**Lung cell line**

BA = -5.5938 - 3.0701\*Mor 22u + 16.242\*MATS1e + 63.408\*HATS8p ... (1)

n=12, R<sup>2</sup>=0.9338; F=37.61; s<sup>2</sup>=0.0530; cross validated R<sup>2</sup> (q<sup>2</sup>)=0.8197

**Breast cell line**

BA = -5.2806 - 2.9754\*Mor 22u + 15.734\*MATS1e + 60.569\*HATS8p ... (2)

n=12, R<sup>2</sup>=0.9310; F=35.97; s<sup>2</sup>=0.0525; cross validated R<sup>2</sup> (q<sup>2</sup>)=0.8083

**CNS cell line**

BA = -11.062 + 96.026\*Rel. No. of O atoms + 60.163\*Rel positive charge (QMPOS/QTPLUS) + 4.4583\*GATS4e ... (3)

n=12, R<sup>2</sup>=0.9692; F=83.96; s<sup>2</sup>=0.0273; cross validated R<sup>2</sup> (q<sup>2</sup>)=0.9264

The expansions of all the descriptors, which were found to be significant in the MLRA, are given in **Table III**.

The compounds **15-18, 21, 22, 25-30** generally demonstrated to have anti-tumor activity rationalizing

the aim of synthesizing steroidal molecule as anticancer agents. To further strengthen the future approaches, QSAR models were developed with the help of these synthesized and tested compounds.

The correlations among the descriptors showing higher cross-validated  $R^2$  values were used to derive the conventional QSAR equation (eqn. 1-3),  $R^2$  and  $s^2$  values. For various cell lines, following descriptors are found to be the best:

In the anticancer study against the lung and breast cell line QSAR models showed dependence mainly upon electrostatic, Whim and Topological descriptors

15            30            0            0            0  
along with small dependence on 3D and 2D autocorrelations. The  $R^2$  values ranges from 0.9338-0.8225 and  $q^2$  values from 0.8197-0.7188 for lung cell line, while the range for  $R^2$  and  $q^2$  values for breast cell lines were 0.9310- 0.8271 and 0.8083-0.7121 respectively.

A positive linear correlation has been found among electrostatic, WHIM and 3D parameters and  $r^2$  ranges from 0.9692-0.9128 and  $q^2$  between 0.9264-0.8570 when anticancer activity was tested against CNS cell line.

To further validate the models developed in QSAR studies, we had run a test set **10**, **18** and **25** using these models and calculated their predicted activity; the results of which are shown in **Table IV**.

The tests set results from **Table III** suggest that the predictive ability of the models is reasonably well when the 3 and 17-position of the steroid ring is substituted by hydroxyl, acetate, pyrrolidino or oximino moieties. Thus, it can be concluded that the focused attention on accurately assessing each step has given rise to the development of reliable QSAR models 3,17-substituted androstene derivatives acting as anticancer agents. The activities of the compounds having a group other than keto and /or enol at 3 and 17 position of steroid nucleus can be predicted satisfactorily using these models.

### Experimental Section

All the aldehydes and aluminium isopropoxide were procured from Fluka Chemie, (Switzerland).

**Table II** — Structures of the compounds selected for *in vitro* anticancer activity along with the results of the activity

| S.No | Compd     | Activity (in growth percentage, conc. used =1.00E-04 M) |        |     |
|------|-----------|---|--------|-----|
|      |           | Lung  | Breast | CNS |
| 1    | <b>10</b> | 97  | 99     | 97  |
| 2    | <b>11</b> | 98  | 95     | 102 |
| 3    | <b>13</b> | 85  | 69     | 101 |
| 4    | <b>15</b> | 0   | 0      | 0   |
| 5    | <b>16</b> | 0   | 0      | -1  |
| 6    | <b>17</b> | 0   | 0      | 0   |
| 7    | <b>18</b> | 0   | 0      | 0   |
| 8    | <b>21</b> | 0   | 0      | 0   |
| 9    | <b>22</b> | 0   | 0      | 13  |
| 10   | <b>25</b> | 0   | 0      | 0   |
| 11   | <b>26</b> | 0   | 0      | 0   |
| 12   | <b>27</b> | 0   | 0      | 0   |
| 13   | <b>28</b> | 0   | 0      | 0   |
| 14   | <b>29</b> | 0   | 0      | 0   |

**Table III** — The expanded version of descriptors as stated in eqn. 1,2 and 3

| S.No. | Descriptor | Description   |
|-------|------------|---|
| 6     | MOR22u     | 3D-MoRSE – signal 22/unweighted (3D-MoRSE descriptors)                        |
| 7     | MATS 1e    | Moran autocorrelation lag1/weighted by atomic Sanderson electronegativities   |
| 8     | HATS8p     | Leverage – weighted autocorrelation of lag 8/unweighted (GETAWAY descriptors) |
| 12    | GATS 4e    | Geary autocorrelation - lag 4/weighted by atomic masses (2D autocorrelation)  |

**Table IV** — The predicted and experimental activity values of the test set

| S. No. | Compd     | Predicted values of anticancer activity (log (% inhibition)) |        |        | Experimental values of anticancer activity (log (% inhibition)) |        |       |
|--------|-----------|--|--------|--------|---|--------|-------|
|        |           | Lungs  | Breast | CNS    | Lungs   | Breast | CNS   |
| 1      | <b>10</b> | 0.4332   | 0.0443 | 0.4752 | 0.030   | 0.040  | 0.050 |
| 2      | <b>18</b> | 1.6651   | 1.8711 | 2.4266 | 2.000   | 2.000  | 2.000 |
| 3      | <b>25</b> | 1.8165   | 1.8594 | 2.0000 | 2.000   | 2.000  | 2.000 |

$^1\text{H}$  NMR spectra were recorded on AC-300F, 300 MHz, Varian EM-390, 90 MHz and EM-360, 60 MHz NMR instruments (Bruker, Switzerland) using tetramethylsilane (TMS) as the internal standard (chemical shifts in  $\delta$ , ppm). IR and UV-Vis spectra were recorded on Perkin-Elmer 882 (Perkin-Elmer, Ltd., England) and Lambda 15 spectrophotometer (Perkin-Elmer, Germany) models respectively. Elemental analyses were carried out on a Perkin-Elmer-2400 (Perkin-Elmer, USA) and the results were within  $\pm 0.4\%$  of theoretical values for C, H and N. Mass spectra were recorded on a V6-11-250J70 S and CEC-21-110B Finnigan Mat 1210 (US) or Micro Mass 7070 (UK) at 70 eV using a direct inlet system.

Software used for QSAR analysis were: Chem3D, Dragon and Codesssa.

**General procedure for synthesis of 16-[4-(substituted)-3-methoxybenzylidene]-17-oxo-5-androsten-3 $\beta$ -ols 10 and 21**

A mixture of dehydroepiandrosterone **9** (1.0 g, 3.46 mM), aldehyde (1.5 g) and sodium hydroxide (1.5 g) in methanol (10 mL) was stirred for 1.5 hr at RT (aldol condensation). The reaction-mixture was then added to ice-cold water. The precipitate obtained was filtered, washed with water and recrystallized from methanol.

**16-[4-(2-Dimethylaminoethoxy)-3-methoxybenzylidene]-17-oxo-5-androsten-3 $\beta$ -ol 10.** A mixture of **9** (1.0 g), 4-(2-dimethylaminoethoxy)-3-methoxybenzaldehyde (**7**, 1.5 mL, 5.03 mM) afforded **10** (1.2 g).

**16-[3-Methoxy-4-(2-morpholin-4-yl-ethoxy)benzylidene]-17-oxo-5-androsten-3 $\beta$ -ol 21.** Aldol condensation between **9** (1.0 g), 3-methoxy-4-(2-morpholin-4-yl-ethoxy)benzaldehyde (**8**, 1.5 g, 4.0 mM) yielded **21** (1.20 g).

**General procedure for synthesis of substituted androstene-3,17-diones 11 and 22**

The aldol product was subjected to Oppenaur oxidation by dissolving in a mixture of cyclohexanone (10 mL) and dry toluene (150 mL). The residue obtained was allowed to stand overnight and then it was filtered, washed with the water, dried and purified by recrystallization from methanol.

**General procedure for synthesis of 16-substituted-3-pyrrolidino-3,5-androstadien-17-ones 12 and 23**

Freshly distilled pyrrolidine (1.0 mL) was added to a refluxing solution of dione **11** and **22**, (0.5 g,

2.0 mM) in methanol (100 mL). Refluxing was continued for 15 min. On cooling in ice the crystalline material obtained was filtered, washed with methanol and dried to afford **12** and **23** respectively, which was then immediately used for next step of synthesis.

**General procedure for synthesis of 16-substituted-3 $\beta$ -pyrrolidino-5-androsten-17 $\beta$ -ol 13, 24, and substituted-5-androstene-3 $\beta$ ,17 $\beta$ -diols 19, 30**

Sodium borohydride (1.0 g, 27.0 mM) was added in small quantities to a stirred suspension of reactants (0.5 g) **12**, **23**, **10** and **21** in methanol (50 mL) at RT. Stirring was continued for further 4 hr. Excess of methanol was removed by distillation under reduced pressure until a small volume of reaction-mixture was obtained, which was poured into 50 mL of ice-cold water. The precipitated product was filtered, washed with water, dried and purified by recrystallization from methanol/ acetone.

**General procedure for synthesis of 16-substituted-3 $\beta$ -pyrrolidino-5-androsten-17 $\beta$ -yl acetates 14,17, 20,25, 28 and 31**

The acetylation of the reactants (0.5 g) **13**, **10**, **19**, **24**, **21** and **30** was performed in dry pyridine (1.5 mL) using acetic anhydride (1.0 mL).

**Pharmacological Activity**

The synthesized compounds were sent to National Cancer Institute, Bethesda, USA. The selected compounds were evaluated for anticancer activity at 0.1 mM conc. using 3-cell lines panel consisting of MCF-7 (breast), NCI-H460 (lung) and SF-268 (CNS). End-point determinations were made with sulforhodamine B, a protein binding dye. Results of each test agent were reported as the percent of growth of treated cells when compared to the untreated control cells. Compounds that reduced the growth of any one of the cell lines to 32% or less (negative numbers indicate cell kill) were termed as active.

**Quantitative Structure Activity Relationship Study**

The statistical parameters used during the development and validation of QSAR models will be discussed, which are listed below.  $R^2$  is the correlation coefficient,  $q^2$  is the leave-one-out cross-validated  $R^2$ ;  $F$  and  $s^2$  are the  $F$  value and the standard deviation of the regression, respectively.

**Data set preparation:** Cronin and Shultz<sup>18</sup> suggested that highly homogeneous biological data

are required to develop a QSAR model with good predictive capability. High quality biological data are ideally measured by the same protocol, within the same laboratory and by the same operator. Thus the data of steroid derivatives fall in the above-mentioned criterion exactly. The wide range of activity of these compounds that is due to the various substituents attached makes this series suitable for a QSAR investigation. The log of biological activity was taken for QSAR analysis. The initial geometries of the compounds were built using Chem Draw software; the energy of all the molecules was minimized by molecular mechanics- MM2 force field. The dataset was split into a training set (TR) and a test set (TS) using a stratified random selection method. The TS included 3 compounds, whose activity reflected the whole range of activities of the TR, to get the TS as much representative as possible of the whole dataset.

#### Data analysis and QSAR equation development

A total of about 1700 different molecular descriptors grouped into different families: physico-chemical, topological, electro-topological, geometrical, constitutional, 2D and 3D descriptors were calculated by the software- Chem3D, Dragon and Codessa. Constitutional and topological descriptors require only 2D molecular structure information. The geometrical descriptors, instead, are based on the 3D coordinates; electrostatic ones reflect charge distribution of the molecules. The protocol used to select the best descriptor models is summarized below.

(i) For each class, correlation equation with 1, 2... $ND_{max}$  descriptors (maximum number of descriptors) were calculated using the heuristic correlation, which automatically accomplished by the program. The heuristic pre-selection method, accurately described in the CODESSA reference manual, is based on a stepwise procedure, which first of all, discards all the descriptors with issuing values or poor degree of variation between different structures. Further, highly inter-correlated descriptors are also excluded. At this level, the use of some control parameters for heuristics pre-selection of descriptors is required to avoid over-correlation. A QSAR model is valid and stable when ratio of the number of observations (compounds) and the number of variables (descriptors) is at least 4:1. The control parameters were set at little lower limits than those taken by default by the

program, to decrease further the acceptable inter-correlation allowed between the descriptors. Maximum number of descriptors,  $ND_{max}=3$ ; one parameter significance criteria,  $R^2_{min}=0.01$ ; high inter-correlation level,  $r_{full}=0.99$ ; significant inter-correlation level,  $r_{sig}=0.8$ ; one parameter t-test for significance,  $t_1=0.1$ ; multi-parameter t test for significance,  $t_2=3$ ; branching criteria,  $NS=3$ .

- (ii) Based on the pre-selected descriptors and the additional descriptors loaded in the file, which were calculated by Dragon software, the program provides the basic correlation equation by performing multi linear regression analysis (MLRA).
- (iii) Each equation contained a different number of descriptors. In general every additional descriptor improves the statistical quality of a model, besides also increasing its complexity. In this work, the selection of the best model was performed, by considering the variations in both  $q^2$  and  $R^2$  of the TR as the number of involved descriptors increased<sup>19</sup>. The point beyond which no significant improvements could be observed is so identified; the corresponding model is selected as the best one in that particular cell line.
- (iv) Descriptors included in the best QSAR equations were used for the prediction of the TS.

**Validation of QSAR model<sup>20</sup>:** One of the most-used criteria for an internal estimate of the predictive ability of QSAR model is  $q^2$ . But, it is also known that high value of  $q^2$  is not sufficient assessing a highly predictive power of the model. To accurately estimate the predictive ability of the model, further conditions over the validation parameters of the test set are required. First of all, QSAR models were submitted to the so-called 'internal validation' (check the  $R^2$  and  $q^2$  over the TR). Only models with  $q^2 > 0.5$  were selected for the so-called 'external validation' (check the  $R^2$  and other parameters over the TR). Such additive validation parameters were analyzed. In the plot of experimental *versus* predicted activities, the obtained regression line ( $y = ax + b$ ) is characterized by a correlation coefficient  $R^2$ . In an ideal QSAR model, the slope of the line (a) is 1, while the intercept with y-axis (b) is 0 and  $R^2$  (varying between 0 and 1) is 1.  $R^2$  calculated on the TS must be  $\geq 0.6$ . When the regression line is forced to pass through the origin of axes (intercept set to 0), its slope (k) should be as close as possible to 1, and close to the slope of the

actual regression line as well.  $0.85 \leq k \leq 1.15$  is suggested to be in an acceptable range. Finally, the correlation coefficient  $R_0^2$  of the regression line forced through the origin should be as close as possible to the value of  $R^2$ , so that  $(R^2 - R_0^2)/R^2 < 0.1$ . In conclusion, we adopted as validation criterion for our QSAR model the one stating that all the following conditions are simultaneously satisfied:  $q^2 > 0.5$ ,  $R^2 > 0.6$ ,  $0.85 \leq k \leq 1.15$ , and  $(R^2 - R_0^2)/R^2 < 0.1$ .

### Acknowledgements

The authors are thankful to the Council of Scientific and Industrial Research, New Delhi, India, for providing the financial support and NCI, USA for carrying out the anti cancer screening of the compounds.

### References

- 1 Vogel C L, *Anticancer Drugs*, 14, **2003**, 265.
- 2 Bedir E, Khan I A & Walker L A, *Pharmazie*, 57, **2002**, 491.
- 3 Kuo K W, Hsu S H, Li Y P, Lin W L, Liu L F, Chang L C, Lin C N & Sheu H M, *Biochem Pharmacol*, 60, **2000**, 1865.
- 4 Luo L, Li Z, Zhang Y & Huang R, *Yao Xue Xue Bao*, 33, **1998**, 839.
- 5 Brodie A M H, Coombes R C & Dowsett M, *J Steroid Biochem*, 27, **1987**, 899.
- 6 Marsh D A, Brodie H J, Garrett W, Tsai-Morris C H & Brodie A M, *J Med Chem*, 28, **1985**, 788.
- 7 Budzar U, *Clin Breast Cancer*, 4, **2003**, S84.
- 8 Dixon J M, *Expert Opin Pharmacother*, 5, **2004**, 307.
- 9 Goss P E & Strasser K, *J Clin Oncol*, 19, **2001**, 881.
- 10 Lam H-Y P, Begleiter A & Goldenberg G J, *J Med Chem*, 22, **1979**, 200.
- 11 Chavis C, de Gourey C, Borgna J-L & Imbach J-L, *Steroids*, 39, **1982**, 129.
- 12 Tang W, Scheinder M R & Eisenbrand G, *Anticancer Drug Des*, 13, **1998**, 815.
- 13 Chattopadhyaya R, Ph.D. Thesis, Panjab University, Chandigarh, India, **1999**.
- 14 Christiansen R G, Clinton R O & Dean J W, *US Pat*, 316364, **1964**; *Chem Abstr*, 62, **1965**, 7834.
- 15 Easthem J F & Taranishi R, *Organic Synthesis*, Vol.2, edited by R Adams (John Wiley and Sons Inc, New York), **1960**, pp. 1.
- 16 Dubey S, Piplani P & Jindal D P, *Indian J Chem*, 44B, **2005**, 2126.
- 17 Dubey S, Piplani P & Jindal D P, *Lett Drug Design & Discovery*, 2, **2005**, 537.
- 18 Cronin M T D & Schultz T W, *Theochem*, 39, **2003**, 622.
- 19 Katrizky A R, Fara D C & Karelson M, *Bioorg Med Chem*, 12, **2004**, 3027.
- 20 Golbraikh A & Tropsha A, *J Mol Graphics Modell*, 20, **2002**, 269.