Anticancer activity trends of 5-substituted 2(2-diethylamino) ethyl Anthrapyrazoles toward L1210 murine leukemia: a QSAR Analysis
Howida Abdallah Hashim, Mohamed Osman El-Fakii, and Ahmed Elsadig M. Saeed

ABSTRACT
The anticancer activities of ten known 5-Substituted 2-(2-Diethylamino) ethyl Anthrapyrazoles toward L1210 murine leukemia were picked from literature and correlated with their calculated molar volume and polarizability. In spite of their high activity and identical structure except for C5 side chain, the ten anthrapyrazole failed to give acceptable correlation. A method has to be devised to uncover the underlined cause for this. This was simply done by sequential removal of data points from the data sets and securing the change in the value of statistical correlation coefficient; any points, the removal of which increase the coefficient by circa 0.4 units, was set aside until a maximum value of the coefficient was obtained. This lead to separate five compounds out of ten in a subset with high correlation coefficient (above 0.9). The removal of points was statistically justified by calculating Cook’s squired distance. Regression analysis was performed for the removed points as one set and themselves were found to give high correlation. Thus two distinct sub groups were found to be nested inside the larger group of ten compounds. Statistical and chemical reasoning was exploited to suggest that the two groups are probably mechanistically distinct.

INTRODUCTION
Anthrapyrazoles (1; X=C, R1-R4 various groups) are synthetic anticancer agents that were designed in anticipation of alleviating undesired cardiotoxicity observed in anthracyclene based agents\(^{(1,2)}\) such as anthracyclines (e.g., doxorubicin 2\(^{(3)}\)) and mitoxantrone\(^{(3)}\)\(^{(4)}\). Cardiotoxicity are thought to be caused by the ability of these drugs to generate active oxygen species (ROS) like superoxide and peroxide radicals which attack the mitochondria in heart muscle cells via Fenton chemistry\(^{(5)}\). Anthrapyrazoles could be considered as modification of mitoxantrone in which the central quinone moiety is modified to quasi-iminoquinone in addition to incorporation of a pyrazole ring into the chromophore. This modification was leads to appreciable reduction of cardiotoxicity as was anticipated possibly via alteration of redox cycling associated with their reduced tendency to form semiquinone radicals\(^{(6,7)}\). Like anthracyclines, Anthrapyrazoles are thought to exert their action mainly through intercalation between DNA double strands\(^{(8)}\) and subsequent events which include, among others, interaction with topoisomerase II\(^{(9)}\), DNA strand breaking\(^{(10)}\). From scores of known anthrapyrazoles losoxantrone [1; R1 = (CH\(_2\))\(_2\)N(CH\(_3\))\(_2\), OH, R\(_3\) = H, R\(_4\) = H\(_2\)N\(_2\)] appear to be most promising with a response rate of 63% for breast cancer\(^{(11)}\). Nevertheless, cardiotoxicity is still manifested in. This pave the way to the appearance of yet another related class, the 9-aza-anthrapyrazoles (9-aza-Ap’s) which has a better therapeutic index and a diminished cardiotoxicity\(^{(12)}\). Two promising candidates of 9-aza-Ap’s are BBR 3438 (1; R1 = (CH\(_2\))\(_2\)N(CH\(_3\))\(_2\)OH, R2 = (CH\(_2\))\(_2\)N(CH\(_3\))\(_2\)OH, R3, R4 = H X = N) and BBR 3576 (1; R1 = (CH\(_2\))\(_2\)N(CH\(_3\))\(_2\)OH, R2 = (CH\(_2\))\(_2\)N(CH\(_3\))\(_2\), R3, R4 = H, X = N), the 9-aza-Ap’s are
distinguished by their activity against prostatic carcinoma which show resistance against other anticancer agents. In contrast to anthrapyrazoles, 9-aza-AP’s were shown to have no interaction with topoisomerase II, this suggest a non-enzyme mediated cytotoxic mechanism which might be associated to redox cycling. Quantitative structure activity relationship (QSAR), the methodology by which the present study has been executed, is a category of computational chemistry which endeavor to mathematically correlate the observed biological response to the measured or calculated physicochemical parameters. Although its root went back to the turn of the last century, it did not take momentum until 1960’s, when Hansch attempted to derive a Hammet biological equation inspired by the work of Hammet in the field of physical organic chemistry. QSAR is now well established scientific disciplines utilized in drug design and discovery as well as in toxicology and pharmacology. In this study we attempt to show that the observed activity of this particular group of anthrapyrazoles is a resultant of interplay of more than one factors each acting via physicochemistry of these agents but the relative contribution of each factor may vary in going from one subgroup to another within same main group.

MATERIAL AND METHOD
Design of the molecules and calculations were carried out in a Toshiba laptop. The operating system is Microsoft Vista Ultimate 7, Version 2009, Service Pack 1. Molecules were drawn and their molar volume and polarizability were calculated using ACD/Chemsketch freeware. Anticancer activity data were taken from literature. Statistical analysis were executed using Microsoft Excel 2007. Graphing was performed using Matlab software Version 7 release 14. Chemical Structures of Anthrapyrazoles

Table 1 structures, molar volumes, polarizability and biological activities of 5-Substituted 2-(2-(Diethylamino)ethyl Anthrapyrazoles

<table>
<thead>
<tr>
<th>Compound</th>
<th>R</th>
<th>MV</th>
<th>Pol</th>
<th>pIC50</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-(CH3)2NH2</td>
<td>295.8</td>
<td>43.66</td>
<td>7.49485</td>
</tr>
<tr>
<td>B</td>
<td>-(CH3)2NHCH3</td>
<td>318.9</td>
<td>45.87</td>
<td>7.221849</td>
</tr>
<tr>
<td>C</td>
<td>-(CH3)2NH(CH3)2OH</td>
<td>333.1</td>
<td>48.11</td>
<td>6.69897</td>
</tr>
<tr>
<td>D</td>
<td>-(CH3)2N(CH2CH3)2</td>
<td>372.3</td>
<td>51.64</td>
<td>7.337242</td>
</tr>
<tr>
<td>E</td>
<td>-(CH3)2N(CH2CH3)2</td>
<td>388.3</td>
<td>53.46</td>
<td>7.568636</td>
</tr>
<tr>
<td>F</td>
<td>-(CH3)2N(CH2CH3)2</td>
<td>404.4</td>
<td>55.29</td>
<td>7.49485</td>
</tr>
<tr>
<td>G</td>
<td>-(CH3)2N(CH2CH3)2</td>
<td>452.5</td>
<td>60.77</td>
<td>6.408935</td>
</tr>
<tr>
<td>H</td>
<td>-(CH3)2-c-N(CH2CH3)2O</td>
<td>352</td>
<td>51.32</td>
<td>6.283997</td>
</tr>
<tr>
<td>I</td>
<td>-(CH3)2-c-N(CH2CH3)2NH</td>
<td>350.2</td>
<td>51.98</td>
<td>6.200608</td>
</tr>
<tr>
<td>J</td>
<td>-(CH3)2-c-N(CH2CH3)2NCBZ</td>
<td>459.6</td>
<td>66.48</td>
<td>6.200659</td>
</tr>
</tbody>
</table>

MV= molar volume. Pol= polarizability. pIC50 = the negative logarithm of the inhibitory concentration for 50% of subjects in cultured L1210 murine leukemia cell line. CBZ= benzoyloxycarbonyl group Each column – which represent numerical values of parameter- was correlated with the last column, which represent biological activity in the form of the negative logarithm of the inhibitory concentration for 50% of subjects in cultured L1210 murine leukemia cell line (pIC50). This yields two data sets; data set I in which molar volumes(MV) and pIC50 are paired and data set II in which polarizability (pol) and pIC50 are paired. This was followed by running regression analysis and registering equation and statistical correlation coefficient R² which is equal to +1 when there a perfect positive correlation, -1 when there a perfect negative correlation and it is equal to zero when there is no correlation. Owing to the poor R² values obtained by the above method, data points were sequentially removed from each data set until maximum R² value was attained. The removed points are considered as outliers and their removal was justified by calculating Cook’s squired distance (CD²) and by considering their residual plots. The removed points were themselves put in a group for which regression equation and R² were likewise calculated. R² statistic was used as index to discern the activity trends of anthrapyrazoles

RESULTS & DISCUSSION
The present group of 5-substituted 2-(2-diethylamino)ethyl anthrapyrazoles (compounds a-j: see table 1 for structural details) do not constitute an inherent homogeneous single group regarding the mechanism of their anticancer activity against L1210 murine leukemia cell line. This assumption is based upon the poor correlation (small R² value) which we obtained upon regressing their MV and pol against pIC50 (sets 1 and 2 respectively) as is shown in Table 2 below:
points (d, e, f, g, and j) from each of the two data sets by presence of more than one group exerting different mechanisms and/or spanning different physicochemical domain. Thus a method need to be devised to segregate any mechanistically homogenous sub groups which may exist within the set. Thus we employ the method of sequential removal of data points from sets I and II described above.

After many trials it has been found that removal of data points (d, e, f, g, and j) from each of the two data sets leave the remaining data points (a, b, c, h, and i) as homogenous subsets as indicated by the high $R^2$ values. We designate these as I-A and II-A.

The removed points (d, e, f, g, and j) were themselves arranged in two set which are also homogeneous. We designate these groups as I-B and II-B. The removal of data points d, e, f, g, and j, from sets I and II is statistically justified by calculating Cook’s squared distance ($CD^2$) for them upon sequential replacement of each in data set I-A (or II-A). Cook’s distance is calculated by equation (7) (19) below:

$$CD^2 = \frac{\sum_{i=1}^{n} (\epsilon_i - \hat{\epsilon}_i)^2}{2s^2}$$

All the ten compounds are active against L1210 murine leukemia cell line, the poor correlation between their activities and physicochemical parameters may be caused by presence of more than one group exerting different mechanisms and/or spanning different physicochemical domain. Thus a method need to be devised to segregate any mechanistically homogenous sub groups which may exist within the set. Thus we employ the method of sequential removal of data points from sets I and II described above.

After many trials it has been found that removal of data points (d, e, f, g, and j) from each of the two data sets leave the remaining data points (a, b, c, h, and i) as homogenous subsets as indicated by the high $R^2$ values. We designate these as I-A and II-A.

The removed points (d, e, f, g, and j) were themselves arranged in two set which are also homogeneous. We designate these groups as I-B and II-B. The removal of data points d, e, f, g, and j, from sets I and II is statistically justified by calculating Cook’s squared distance ($CD^2$) for them upon sequential replacement of each in data set I-A (or II-A). Cook’s distance is calculated by equation (7) (19) below:

$$CD^2 = \frac{\sum_{i=1}^{n} (\epsilon_i - \hat{\epsilon}_i)^2}{2s^2}$$

Table 2: QSAR equation and statistical parameters

<table>
<thead>
<tr>
<th>Set</th>
<th>Equation</th>
<th>n</th>
<th>$R^2$</th>
<th>F</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$pIC_{50} = -3 \times 10^{-3} (MV)^2 + 0.02 MV + 3.768$ (1)</td>
<td>a-j (10 points)</td>
<td>0.137</td>
<td>0.001587</td>
<td>0.17854</td>
</tr>
<tr>
<td>I-A</td>
<td>$pIC_{50} = -0.0002 (MV)^2 + 0.107 MV - 6.675$ (2)</td>
<td>a,b,c,h,i (5 points)</td>
<td>0.979</td>
<td>0.935244</td>
<td>0.54312</td>
</tr>
<tr>
<td>I-B</td>
<td>$pIC_{50} = -0.00033 (MV)^2 + 0.263 MV - 44.50$ (3)</td>
<td>d,e,f,g,j (5 points)</td>
<td>0.996</td>
<td>0.803374</td>
<td>0.56593</td>
</tr>
<tr>
<td>II</td>
<td>$pIC_{50} = -0.00077 (pol)^2 + 0.042pol + 6.81$ (4)</td>
<td>a-j (10 points)</td>
<td>0.238</td>
<td>0.01859</td>
<td>1.36763</td>
</tr>
<tr>
<td>II-A</td>
<td>$pIC_{50} = 0.0027(pol)^2 - 0.423pol + 20.76$ (5)</td>
<td>a,b,c,h,i (5 points)</td>
<td>0.989</td>
<td>0.96592</td>
<td>0.58014</td>
</tr>
<tr>
<td>II-B</td>
<td>$pIC_{50} = -0.00035 (pol)^2 - 0.057pol - 11.45$ (6)</td>
<td>d,e,f,g,j (5 points)</td>
<td>0.854</td>
<td>0.87985</td>
<td>0.59624</td>
</tr>
</tbody>
</table>

$n =$ number of data points, $R^2 =$ statistical correlation coefficient, $F =$ Fischer statistic, $SD =$ standard deviation

$\hat{\epsilon}$ is a predicted $y$-value obtained when all the data points are used and

$\hat{\epsilon}_i$ is the corresponding predicted $y$-value obtained when the $i$th point is omitted: $2s^2/n$ is a statistic which estimate the random error in the $y$-direction. Values of $CD^2$ greater than 1 justify the omission of the suspect point. $CD^2$ for the removed data points (which constitute subset I-B) are shown in Table (3) below:

Table 3 CD2 test for the removed data points

<table>
<thead>
<tr>
<th>Data point</th>
<th>CD2</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>0.7766</td>
</tr>
<tr>
<td>e</td>
<td>1.41</td>
</tr>
<tr>
<td>f</td>
<td>1.25</td>
</tr>
<tr>
<td>g</td>
<td>1.676</td>
</tr>
<tr>
<td>j</td>
<td>1.787</td>
</tr>
</tbody>
</table>

As it is clear from Table (3) all the removed data points give $CD^2$ values greater than 1 except point d which gives a value less than 1.0. Nevertheless residual plot (Figure 1) of set I-A including point d shows that all points of set I-A distribute within ± 0.5 of $pIC_{50}$ units while point d lie further away (about ± 0.75 unit). This indicate that this point is unmistakably an outlier. Furthermore, $R^2$ decreases from 0.979 to 0.659 upon replacing point d back into set I-A.
So far we have been discussing statistical justification of separating data sets I and II into two subsets each. Now we consider physicochemical justification of such a separation. Upon scrutinizing the chemical structures of compounds in subsets I-A and I-I A (Scheme 1) we could discern a clear similarity between all of them.

This similarity is represented by the presence of a two-carbon spacer chain between the proximal and distal amino groups of the side chain. The distal amino groups are primary for compound (a) secondary for compound (b) and (c) and part of saturated heterocyclic rings for the compound (h) and (i). The activities of groups I-A and II- A directly correlate with both molar volume and polarizability respectively with slight preference of the latter ($R^2 = 0.989$) over the former ($R^2 = 0.979$). This may has to do with the presence of sterically unhindered lone pairs of electrons on the distal amino nitrogen in all the compounds of this group. These lone pairs have also a direct impact on the basicity for subset I-A which could be deduced from the basicity of the parent amines from which these side chains are derived. The basicity of the parent amines is as follows: methylamine ($pK_a = 10.62$) > dimethylamine($pK_a = 10.77$) > (2-methylamino) ethanol ($pK_a = 9.88$) > N-methylpiprazine($pK_a = 9.09$) > N-methylmorpholine ($pK_a = 7.13$).

The above order indicate that basicity of the side chain (and, indeed, that of the whole molecule) is proportional to activity [see Figure 2]. As indicated by Figure 2, dimethylamine [correspond for compound (b)] have highest basicity due to electron donating effect of the peripheral methyl group which does not apply for methylamine [correspond for compound (a)]. the lowering of basicity in going from compound (c) through compound (h) to compound (i) is quite consistence with the presence in each of a second hetero atom which withdraw electron density from the middle amino group making the lone pair less available.

Figure (1) : Regression line and residual plot obtained by replacing data point d in subset I-A

Scheme 1: Compounds of subsets I-A and I-B
distribution of points along the curve in Figure 2 embody the parabolic trend almost perfectly ($R^2 \approx 0.93$).

Groups I-B and II-B (Scheme 2), on the other hand, are similar in that terminal group in the side chain at C5 is a dithylamino group except compound (j) in which the side chain at C5 end up with a piprazine moiety protected at distal N4’ with by benzylxycarbonyl (CBZ) group. We expect tertiary amino groups have somewhat similar basicities. This is also applies for compound (j) in which N4’ is masked from affecting N1’ by CBZ group making it similar to its counterparts compounds d – g. Group I-B compounds also exhibit another similarity in that the peripheral N atoms in C5 side chain are all surrounded by bulky group as is shown by Scheme 2 below. Basicity is roughly the same throughout subsets I-B and II-B while the variation of biological activity is a function of other physicochemical parameters. This is in contrast to set I-A and II-A where activity correlate with basicity. The activities of groups I-B and II-B also parabolically correlate with both molar volume and polarizability respectively. This time the preference is for of the former ($R^2 = 0.996$) over the latter ($R^2 = 0.854$). This is indicative that molar volume play a more important role in determining the mechanism of action of this particular series. Indeed the compounds of these group are bulkier compared to those of groups A, which implies that the mere intercalation of these groups into DNA strands may create a larger cavity. As it is well known for all intercalating anticancer agents, the creation of such cavity necessitate that DNA undergo a conformational change involving an increase in the vertical separation between the base pairs and consequently leading to partial unwinding of the double helix. This, in turn leads to distortion of sugar-phosphate backbone and alteration in twist angle between the base pair. Thus, these effects might be more pronounced in subsets B.

Figure 3 demonstrates the curves results from plotting of the correlations traced by subsets I and II for molar volume (i) and polarizability (ii). In this figure, each of the two groups of compounds trace a different curve line for each of the two physicochemical parameters employed. Note the clear segregation of their molar volume and polarizability domains while the range of biological activity are the same. For polarizability, the separation of the two curves for subset I-A and II-B are somewhat narrower.

Figure 2. The correlation of the basicities of terminal amino groups (pKa’s) with biological activities (pIC$_{50}$’s) of subsets I-A (or II-A)

Scheme 2 Compounds of subsets II-A and II-B
It is well known that activity of anticancer agent are multifactorial\(^3\). It is unlikely that a simple model such as the present one could convey the intractable mode of action of the various anthrapyrazoles. Suffice it to say that the authors attempt to explain the behavior of the limited number \(-\)substituted \(2\)-(diethylamino) ethyl anthrapyrazoles. The sites of variability of the broader class of anthrapyrazoles are exclusively three; the chromophore, N2 side chain and C5 side chain. Research should be directed toward quantifying and exposing the impact of variation the side chains at N2 and C5 for a fixed chromophore structure. The problem could also be tackled the other way round by fixing the side chains and varying the chromophore. This may yield interesting outcomes in the quest for better therapeutic anthrapyrazoles.

CONCLUSION

In such a small group as the ten anthrapyrazole used in the present study, a variation of mode of action could exist in spite of the fact that the compounds are identical except for the C5 side chain. The observed poor correlation between physicochemical parameters and biological activity results from presence of two subgroups created by the variation of C5 substituents and nested within the larger group. By sequential removal of data points and observing the change in \(R^2\), we have succeeded to separate the nested two group and give reasonable justification for this separation both statistically and chemically.

REFERENCES


17. ACD/Labs release 12.00 product version 12.01(2012), available at: www.acdlabs.com


