Dysregulation of cell-cycle leading to an uncontrolled cellular proliferation is a universal characteristic of cancer (Sherr, 2000). Progression of cells through the cell-cycle is dependent on the formation of specific protein kinase complexes called as cyclin dependent kinases (CDKs) which form in a cyclical fashion (Yeudall and Jakus, 1995). The CDKs are a family of heterodimeric Ser/Thr protein kinases each consisting of a catalytic CDK subunit and an activating cyclin subunit (Pavletich, 1999; Malumbres and Barbacid, 2005). Activities of CDKs are controlled by association with cyclins and reversible phosphorylation reactions (Hengstschl’ager et al., 1999). The association of the CDKs with their requisite cyclin partner results in the CDKs adopting a substrate-specific catalytic subunit (Sridhar et al., 2006). The biological activity of CDKs is negatively controlled by direct interactions with proteins referred to as CDK inhibitors. CDK inhibitors are divided into two major families: the Ink4 family, which specifically inhibit cyclin D-associated kinases (CDK 4 and 6) and the Cip/Kip family which inhibit most of the CDKs (Zieske, 2000). The CDK4 initiate the functional change from quiescence (G_0) to proliferation. The major Cdk4/cyclin D substrate is the product of the retinoblastoma gene (pRB) (Morgan, 1997). During G1, pRB induces the members of a family of cell cycle regulatory transcription factors, collectively referred to as E2Fs, to activate the transcription of genes whose products are required for S phase (Lee et al., 1987; Julie et al., 2005). Alterations in the cascade involving CDK4, CDK6, cyclin D, Ink4, pRB and E2F have been observed in more than 80% of human cancers.

Therefore, the development of selective CDK4 inhibitors is a promising approach for cancer therapy (Dai and Grant, 2003; Shapiro, 2006; Thomas et al., 2007; Graf et al., 2009). In addition, the identification of specific amino acid residues of the kinase superfamily around the ATP-binding pocket of CDK4 have enabled the researchers to develop potent and selective CDK4 inhibitors (Honma et al., 2001). The structure-based design of potent and selective CDK4 inhibitors led to the development of several classes of compounds, including pyrido[2,3-d] pyrimidines (Toogood et al., 2005), 2-anilinopyrimidines, diaryl ureas, benzoyl-2,4-diaminothiazoles, indolo[6,7-a] carbazoles (Zhu et al., 2003), pyrrolo[3,4-c]carbazoles (Engler et al., 2003; Sanchez-Martinez et al., 2003) and oxindoles (Sharma et al., 2008).
The various CDK inhibitors that are currently in preclinical and clinical trials are UCN-01, PD 0183812, Flavopiridol (Grant and Roberts, 2003), R547 (Chu et al., 2006), AT7519 (Wyatt et al., 2008), SNS-032, Roscovitine (CYC202) JNJ-7706621, AG-024322, AT7519, AZD5438, P1446A-05, P276-00 and PD-0332991. Among the above mentioned the compound PD 0332991 and P1446A-05 are selective CDK4 inhibitors (McInnes, 2008; Lapenna and Giordano, 2009). Despite more than a decade of investigation, none of the CDK inhibitors resulted in drug approval owing to low activity and toxicity in the clinical trials (Shapiro, 2006; Lapenna and Giordano, 2009). Consequently, there is a strong need to develop potent and selective CDK inhibitors.

In the present study both classification and correlation approaches have been successfully employed for development of models for the prediction of CDK4 inhibitory activity of aminomethylene isoquinoline-1,3(2H,4H)-diones (Tsou et al., 2008).

**Model Development**

All the 52 4-aminomethylene-4-$H$ isoquinoline-1,3-dione derivatives reported by Tsou et al. as CDK4 inhibitors were selected as a dataset for the purpose of present study (Tsou et al., 2008). The basic structures for the said derivatives are shown in Figure 8.1. and various substituents enlisted in Table 8.1. MDs of diverse nature were used in the study. These included physico-chemical descriptors, path count, path cluster, estate contribution descriptors, polar surface area descriptors, element counts, topological descriptors and a variety of alignment independent descriptors. All computational work was performed on Apple workstation (8-core processor) using V-life MDS QSAR plus developed by V life sciences technologies Pvt. Ltd, Pune, India. The values of other MDs which are not the part of V-life MDS QSAR plus were computed using an in-house computer program. MDs with significant degenerate values were omitted from a large pool of descriptors Finally, 46 descriptors (enlisted in Table 8.2) were shortlisted on the basis of non-correlating nature and classification ability and subsequently employed for present study are as per following:
SssOE-index (Estate Contribution)
This is an electrotopological state index for number of oxygen atom connected with two single bonds.

\[ S_{ssOE_i} = I_i + \Delta I_i \]

Where \( S_i \) is the E-state for atom \( i \), \( I_i \) is an intrinsic value and \( \Delta I_i \) is the influence of other atom on \( i \) (Hall et al., 1991; Hall and Kier, 1995).

![Figure 8.1. (A-D) Basic structures and arbitrary atom numbering scheme for the 4-aminomethylene isoquinoline-1,3-dione (Tsou et al., 2008)](image)

SsoHE-index (Estate Contribution)
This is an electrotopological state index for number of hydroxyl (–OH) group connected with one single bond (Hall et al., 1991; Hall and Kier, 1995).
Table 8.1. Relationship of E-state contribution index (SssOE), augmented eccentric connectivity topochemical index ($\chi^A$), molecular connectivity index ($\chi^A$) and connective eccentricity topochemical index ($C^\xi_C$) with CDK4 inhibitory activity

<table>
<thead>
<tr>
<th>Cpd No.</th>
<th>R</th>
<th>A24</th>
<th>A10</th>
<th>A8</th>
<th>A12</th>
<th>Predicted using MAA models</th>
<th>Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>H</td>
<td>24.16</td>
<td>18.71</td>
<td>12.63</td>
<td>5.27</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A2</td>
<td>7-Br</td>
<td>24.43</td>
<td>18.04</td>
<td>12.45</td>
<td>4.66</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A3</td>
<td>6-Br</td>
<td>24.41</td>
<td>18.56</td>
<td>12.45</td>
<td>4.84</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A4</td>
<td>6-NO$_2$</td>
<td>24.4</td>
<td>19.78</td>
<td>13.69</td>
<td>5.48</td>
<td>- ±</td>
<td>±</td>
</tr>
<tr>
<td>A5</td>
<td>6-I</td>
<td>24.42</td>
<td>18.18</td>
<td>12.31</td>
<td>4.53</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A6</td>
<td>O$\overline{\text{C}}$-N(Me)$_2$</td>
<td>37.16</td>
<td>19.94</td>
<td>14.67</td>
<td>5.56</td>
<td>- ±</td>
<td>± ± -</td>
</tr>
<tr>
<td>A7</td>
<td>$\text{N}$</td>
<td>24.94</td>
<td>20.61</td>
<td>14.99</td>
<td>5.65</td>
<td>+</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>A8</td>
<td>$\text{N}$-CH$_3$</td>
<td>36.02</td>
<td>19.20</td>
<td>14.27</td>
<td>5.36</td>
<td>-</td>
<td>- - - -</td>
</tr>
<tr>
<td>A9</td>
<td>$\text{C}$</td>
<td>25.11</td>
<td>20.26</td>
<td>15.49</td>
<td>5.57</td>
<td>- ±</td>
<td>± - ± -</td>
</tr>
<tr>
<td>A10</td>
<td>$\text{O}$</td>
<td>25.06</td>
<td>20.10</td>
<td>15.71</td>
<td>5.59</td>
<td>± ±</td>
<td>- ± - -</td>
</tr>
</tbody>
</table>

Models for the prediction of CDK4 inhibitory activity
Table 8.1. Continued

<table>
<thead>
<tr>
<th>Cpd No.</th>
<th>R</th>
<th>A24</th>
<th>A10</th>
<th>A8</th>
<th>A12</th>
<th>Predicted using MAA models</th>
<th>Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A24</td>
<td>A10</td>
</tr>
<tr>
<td>A_{11}</td>
<td>-</td>
<td>24.91</td>
<td>20.51</td>
<td>14.96</td>
<td>5.63</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A_{12}</td>
<td>-</td>
<td>24.92</td>
<td>20.51</td>
<td>14.96</td>
<td>5.63</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A_{13}</td>
<td>-</td>
<td>25.09</td>
<td>20.98</td>
<td>14.60</td>
<td>5.55</td>
<td>±</td>
<td>+</td>
</tr>
<tr>
<td>A_{14}</td>
<td>-</td>
<td>36.11</td>
<td>19.67</td>
<td>16.45</td>
<td>5.42</td>
<td>-</td>
<td>±</td>
</tr>
<tr>
<td>A_{15}</td>
<td>-</td>
<td>25.11</td>
<td>19.05</td>
<td>16.58</td>
<td>5.42</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A_{16}</td>
<td>-</td>
<td>25.07</td>
<td>19.05</td>
<td>16.58</td>
<td>5.42</td>
<td>±</td>
<td>-</td>
</tr>
<tr>
<td>B_{1}</td>
<td>-</td>
<td>24.15</td>
<td>17.97</td>
<td>12.83</td>
<td>5.24</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B_{2}</td>
<td>-Br</td>
<td>24.4</td>
<td>18.10</td>
<td>12.65</td>
<td>4.83</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B_{3}</td>
<td>-Cl</td>
<td>24.4</td>
<td>18.34</td>
<td>12.92</td>
<td>5.18</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B_{4}</td>
<td>-OMe</td>
<td>24.47</td>
<td>18.51</td>
<td>13.61</td>
<td>5.34</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B_{5}</td>
<td>-NO_{2}</td>
<td>46.56</td>
<td>19.12</td>
<td>13.89</td>
<td>5.46</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 8.1. Continued

<table>
<thead>
<tr>
<th>Cpd No.</th>
<th>R</th>
<th>A24</th>
<th>A10</th>
<th>A8</th>
<th>A12</th>
<th>Predicted using MAA models</th>
<th>Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A24</td>
<td>A10</td>
<td>A8</td>
<td>A12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B_6</td>
<td>-I</td>
<td>24.41</td>
<td>17.77</td>
<td>12.51</td>
<td>4.51</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B_7</td>
<td>N(Me)_2</td>
<td>37.15</td>
<td>19.34</td>
<td>14.87</td>
<td>5.54</td>
<td>-</td>
<td>± ± ± -</td>
</tr>
<tr>
<td>B_8</td>
<td>N</td>
<td>24.93</td>
<td>20.01</td>
<td>15.2</td>
<td>5.63</td>
<td>+</td>
<td>± + + +</td>
</tr>
<tr>
<td>B_9</td>
<td>O(CH)</td>
<td>36.01</td>
<td>18.60</td>
<td>14.47</td>
<td>5.34</td>
<td>-</td>
<td>± - - -</td>
</tr>
<tr>
<td>B_10</td>
<td>N</td>
<td>25.1</td>
<td>19.72</td>
<td>15.7</td>
<td>5.57</td>
<td>-</td>
<td>± - ± -</td>
</tr>
<tr>
<td>B_11</td>
<td>N</td>
<td>25.1</td>
<td>19.93</td>
<td>15.53</td>
<td>5.55</td>
<td>-</td>
<td>± - ± -</td>
</tr>
<tr>
<td>B_12</td>
<td>N+Me</td>
<td>25.17</td>
<td>19.67</td>
<td>15.98</td>
<td>5.50</td>
<td>-</td>
<td>± - ± -</td>
</tr>
<tr>
<td>B_13</td>
<td>Me</td>
<td>25.05</td>
<td>19.02</td>
<td>16.22</td>
<td>5.47</td>
<td>±</td>
<td>- ± ± -</td>
</tr>
<tr>
<td>B_14</td>
<td></td>
<td>25.05</td>
<td>19.53</td>
<td>15.80</td>
<td>5.57</td>
<td>±</td>
<td>± ± ± -</td>
</tr>
<tr>
<td>B_15</td>
<td>O</td>
<td>24.91</td>
<td>19.92</td>
<td>15.16</td>
<td>5.61</td>
<td>+</td>
<td>± ± + +</td>
</tr>
</tbody>
</table>
Table 8.1. Continued

<table>
<thead>
<tr>
<th>Cpd No.</th>
<th>R</th>
<th>A24</th>
<th>A10</th>
<th>A8</th>
<th>A12</th>
<th>Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_{16}</td>
<td></td>
<td>25.02</td>
<td>19.65</td>
<td>15.72</td>
<td>5.59</td>
<td>+ ± - ± +</td>
</tr>
<tr>
<td>B_{17}</td>
<td></td>
<td>24.98</td>
<td>20.46</td>
<td>14.80</td>
<td>5.54</td>
<td>+ - ± ± +</td>
</tr>
<tr>
<td>B_{18}</td>
<td></td>
<td>25.04</td>
<td>19.29</td>
<td>16.07</td>
<td>5.44</td>
<td>± ± - ± -</td>
</tr>
<tr>
<td>B_{19}</td>
<td></td>
<td>25.15</td>
<td>19.09</td>
<td>15.88</td>
<td>5.32</td>
<td>- - - - -</td>
</tr>
<tr>
<td>B_{20}</td>
<td></td>
<td>25.08</td>
<td>19.93</td>
<td>16.12</td>
<td>5.64</td>
<td>± ± ± + +</td>
</tr>
<tr>
<td>B_{21}</td>
<td></td>
<td>25.07</td>
<td>19.34</td>
<td>16.12</td>
<td>5.47</td>
<td>± ± ± ± +</td>
</tr>
<tr>
<td>B_{22}</td>
<td></td>
<td>25.21</td>
<td>19.82</td>
<td>16.57</td>
<td>5.59</td>
<td>- ± ± ± -</td>
</tr>
<tr>
<td>B_{23}</td>
<td></td>
<td>25.18</td>
<td>19.28</td>
<td>16.57</td>
<td>5.44</td>
<td>- ± ± ± +</td>
</tr>
<tr>
<td>B_{24}</td>
<td></td>
<td>25.45</td>
<td>19.64</td>
<td>19.12</td>
<td>5.54</td>
<td>- ± ± ± -</td>
</tr>
<tr>
<td>B_{25}</td>
<td></td>
<td>25.14</td>
<td>18.46</td>
<td>13.72</td>
<td>5.35</td>
<td>- - - - -</td>
</tr>
<tr>
<td>C_{1}</td>
<td></td>
<td>24.49</td>
<td>19.90</td>
<td>12.81</td>
<td>4.70</td>
<td>- ± - - -</td>
</tr>
</tbody>
</table>
Table 8.1. Continued

<table>
<thead>
<tr>
<th>Cpd No.</th>
<th>R</th>
<th>A24</th>
<th>A10</th>
<th>A8</th>
<th>A12</th>
<th>Predicted using MAA models</th>
<th>Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A24</td>
<td>A10</td>
<td>A8</td>
<td>A12</td>
<td>A24</td>
<td>A10</td>
</tr>
<tr>
<td>C2</td>
<td></td>
<td>24.48</td>
<td>19.04</td>
<td>12.75</td>
<td>4.71</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C3</td>
<td></td>
<td>24.53</td>
<td>18.98</td>
<td>13.14</td>
<td>4.70</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C4</td>
<td></td>
<td>24.48</td>
<td>18.55</td>
<td>12.73</td>
<td>4.61</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C5</td>
<td></td>
<td>24.54</td>
<td>19.15</td>
<td>13.14</td>
<td>4.70</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D1</td>
<td></td>
<td>24.89</td>
<td>20.74</td>
<td>14.89</td>
<td>5.66</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D2</td>
<td></td>
<td>24.89</td>
<td>20.75</td>
<td>14.89</td>
<td>5.66</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D3</td>
<td></td>
<td>24.86</td>
<td>20.97</td>
<td>14.83</td>
<td>5.69</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>D4</td>
<td></td>
<td>24.91</td>
<td>22.04</td>
<td>15.26</td>
<td>5.85</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D5</td>
<td></td>
<td>24.87</td>
<td>20.78</td>
<td>14.83</td>
<td>5.63</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D6</td>
<td></td>
<td>24.86</td>
<td>20.76</td>
<td>14.83</td>
<td>5.63</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Note: (+) active compound, (-) inactive compound and (±) compound in transitional range
### Table 8.2. List of molecular descriptors employed for the study

<table>
<thead>
<tr>
<th>Code</th>
<th>Descriptor</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Molecular connectivity topochemical index</td>
<td>Goyal and Madan, 1995; Dureja and Madan, 2005</td>
</tr>
<tr>
<td>A2</td>
<td>Eccentric adjacency topochemical index</td>
<td>Gupta et al., 2003</td>
</tr>
<tr>
<td>A3</td>
<td>Augmented eccentric connectivity topochemical index</td>
<td>Madan and Dureja, 2010</td>
</tr>
<tr>
<td>A4</td>
<td>Eccentric connectivity topochemical index</td>
<td>Kumar et al., 2004</td>
</tr>
<tr>
<td>A5</td>
<td>Connective eccentricity topochemical index</td>
<td>Madan and Dureja, 2010</td>
</tr>
<tr>
<td>A6</td>
<td>Zagreb topochemical index</td>
<td>Bajaj et al., 2005</td>
</tr>
<tr>
<td>A7</td>
<td>Wiener’s topochemical index</td>
<td>Bajaj et al., 2004a</td>
</tr>
<tr>
<td>A8</td>
<td>Molecular connectivity index</td>
<td>Randic, 1975</td>
</tr>
<tr>
<td>A9</td>
<td>Eccentric adjacency index</td>
<td>Gupta et al., 2001</td>
</tr>
<tr>
<td>A10</td>
<td>Augmented eccentric connectivity index</td>
<td>Bajaj et al., 2006</td>
</tr>
<tr>
<td>A11</td>
<td>Eccentric connectivity index</td>
<td>Sharma et al., 1997</td>
</tr>
<tr>
<td>A12</td>
<td>Connective eccentricity index</td>
<td>Gupta et al., 2000</td>
</tr>
<tr>
<td>A13</td>
<td>Zagreb index</td>
<td>Gutman et al., 1975; Gutman and Randic, 1977</td>
</tr>
<tr>
<td>A14</td>
<td>Wiener’s index</td>
<td>Wiener, 1947a; Wiener, 1947b</td>
</tr>
<tr>
<td>A15</td>
<td>Balaban mean square distance index</td>
<td>Balaban, 1983</td>
</tr>
<tr>
<td>A16</td>
<td>Superpendentic index</td>
<td>Gupta et al., 1999</td>
</tr>
<tr>
<td>A17</td>
<td>Rotatable bond count</td>
<td>Todeschini and Consonni, 2009</td>
</tr>
<tr>
<td>A18</td>
<td>S log p</td>
<td>Wildman and Crippen, 1999</td>
</tr>
<tr>
<td>A19</td>
<td>Alignment independent descriptors</td>
<td>Baumann, 2002</td>
</tr>
<tr>
<td>A20</td>
<td>T_C_O_7</td>
<td></td>
</tr>
<tr>
<td>A21</td>
<td>T_2_O_3</td>
<td></td>
</tr>
<tr>
<td>A22</td>
<td>T_2_N_4</td>
<td></td>
</tr>
<tr>
<td>A23</td>
<td>T_C_C_1</td>
<td></td>
</tr>
<tr>
<td>A24</td>
<td>T_2_T_5</td>
<td></td>
</tr>
<tr>
<td>A25</td>
<td>Estate contribution descriptors</td>
<td></td>
</tr>
<tr>
<td>A26</td>
<td>SsSOE-index</td>
<td></td>
</tr>
<tr>
<td>A27</td>
<td>SssCH2E-index</td>
<td></td>
</tr>
<tr>
<td>A28</td>
<td>SdCHE-index</td>
<td></td>
</tr>
<tr>
<td>A29</td>
<td>SaaCHE-index</td>
<td></td>
</tr>
<tr>
<td>A30</td>
<td>Estate contribution descriptors</td>
<td></td>
</tr>
<tr>
<td>A31</td>
<td>SssNHE-index</td>
<td></td>
</tr>
<tr>
<td>A32</td>
<td>SdOE-index</td>
<td></td>
</tr>
<tr>
<td>A33</td>
<td>5PathCount</td>
<td></td>
</tr>
<tr>
<td>A34</td>
<td>ChiV0 chain descriptor</td>
<td>Hall and Kier, 1991</td>
</tr>
<tr>
<td>A35</td>
<td>chiV4 pathCluster</td>
<td></td>
</tr>
<tr>
<td>A36</td>
<td>kappa1</td>
<td>Kier, 1985; Kier, 1986a</td>
</tr>
<tr>
<td>A37</td>
<td>k2alpha</td>
<td>Kier, 1986b</td>
</tr>
<tr>
<td>A38</td>
<td>Polarizability AHP</td>
<td>Miller, 1990a</td>
</tr>
<tr>
<td>A39</td>
<td>Polarizability AHC</td>
<td>Miller, 1990a</td>
</tr>
<tr>
<td>A40</td>
<td>Molecular Weight, MW</td>
<td>Todeschini and Consonni, 2009</td>
</tr>
<tr>
<td>A41</td>
<td>XlogP</td>
<td>Wang et al., 1997</td>
</tr>
<tr>
<td>A42</td>
<td>SMR</td>
<td>Wildman Crippen, 1999</td>
</tr>
<tr>
<td>A43</td>
<td>Hydrogens Count</td>
<td>Todeschini and Consonni, 2009</td>
</tr>
<tr>
<td>A44</td>
<td>Carbons Count</td>
<td>Todeschini and Consonni, 2009</td>
</tr>
<tr>
<td>A45</td>
<td>Polar Surface Area excluding P and S</td>
<td>Palm et al., 1998; Winiwarter et al., 1998</td>
</tr>
<tr>
<td>A46</td>
<td>Superaugmented eccentric connectivity topochemical index 1</td>
<td>Dureja and Madan, 2007</td>
</tr>
<tr>
<td>A47</td>
<td>Superaugmented eccentric connectivity distance sum topochemical index-2</td>
<td>Gupta et al., 2012</td>
</tr>
<tr>
<td>A48</td>
<td>Super adjacency topochemical index</td>
<td>Bajaj et al., 2004b</td>
</tr>
</tbody>
</table>
Alignment independent descriptors

These alignment-independent versatile structure descriptor for QSAR and QSPR based on distribution of molecular features have been defined by Baumann (Baumann, 2002). These include

\[ T_{\text{C}_7} \]

This is the count of number of carbon atoms (having single, double or triple bonds) separated from any oxygen atom (single or double bonded) by 7 bond distances in a molecule (Baumann, 2002).

\[ T_{2\text{O}_3} \]

This is the count of number of double bonded atom from any other Oxygen atom by 3 bonds (Baumann, 2002).

Chi Chain Descriptor

\[ \text{chi5chain} \]

This descriptor signifies a retention index (fifth order) derived directly from gradient retention times for five member ring (Wildman and Crippen, 1999).

Augmented eccentric connectivity topochemical index

It is defined as the summation of the quotients of the product of adjacent vertex chemical degrees and chemical eccentricity of the concerned vertex, for all vertices in a hydrogen suppressed molecular graph (Madan and Dureja, 2010).

\[ \xi^{\text{AC}} = \sum_{i=1}^{n} \left( \frac{M_{ic}}{E_{ic}} \right) \]

Where \( M_{ic} \) is the product of chemical degrees of all the vertices \( (v_j) \), adjacent to vertex \( i \) , \( E_i \) is the chemical eccentricity and \( n \) is the number of vertices in graph \( G \) (Madan and Dureja, 2010).
**Molecular connectivity topochemical index**

It is defined as the summation of the modified bond values of adjacent vertices for all edges in the hydrogen-suppressed molecular graph (Goyal and Madan, 1995; Dureja and Madan, 2005).

\[
\chi^A = \sum_{i=1}^{n} \left( V_i^c V_j^c \right)^{-1/2}
\]

Where \( V_i \) and \( V_j \) represents modified valencies of a pair of vertices joined by edges \((i, j)\) (Goyal and Madan, 1995; Dureja and Madan, 2005).

**Connective eccentricity topochemical index**

It is defined as summation of the ratios of the chemical degree of a vertex and its chemical eccentricity for all vertices in the hydrogen suppressed molecular structure (Madan and Dureja, 2010).

\[
C^c = \sum_{i=1}^{n} \left( \frac{V_{ic}}{E_{ic}} \right)
\]

Where \( V_{ic} \) is chemical degree of vertex \( i \), \( E_{ic} \) is the chemical eccentricity and \( n \) is the number of vertices in graph \( G \) (Madan and Dureja, 2010).

**Eccentric adjacency topochemical index**

It is defined as the summation of the ratios of sum of the chemical degrees of adjacent vertices and eccentricity of the concerned vertex, for all vertices in a hydrogen suppressed molecular structure (Gupta et al., 2003).

\[
\xi^c = \sum_{i=1}^{n} \left( \frac{S_{ic}}{E_i} \right)
\]

Where \( S_{ic} \) is sum of chemical degrees of vertices adjacent to vertex \( i \), \( E_i \) is the eccentricity of vertex \( i \) and \( n \) is the number of vertices in graph \( G \) (Gupta et al., 2003). In present study, R program (version 2.1.0) along with the RPART library was utilized to grow DT. The active compounds were labeled as “A” (n = 15) and the inactive compounds were similarly labeled as “B” (n = 37. Each analogue was assigned a biological activity which was subsequently compared with the reported CDK4 inhibitory activity. The RFs were grown separately for CDK4 inhibitory activity with the R program (version 2.1.0) using the random forest library. For the
purpose of MAA based models the compounds having reported IC$_{50}$ values of $\leq 0.25$ μM were considered to be active (and labelled as “A” ($N = 18$)) while those possessing IC$_{50}$ values $> 0.25$ μM were treated to be inactive (and labelled as “B” ($N = 38$)). The CDK4 inhibitory activity assigned to each compound was subsequently compared with the reported biological activity. The average IC$_{50}$ (μM) values for each range were also calculated.

For correlation models, the dataset was divided into training and test set by random selection method for MLR, PLSR and PCR methods using pIC$_{50}$ [pIC$_{50} = -\log$ (IC$_{50} \times 10^{-6}$)] as dependent variable and various descriptors as independent variables. These models were generated using a training set of 36 molecules. Predictive power of the resulting models was evaluated by test set of 16 molecules. The biological activities of all the compounds had uniform distribution ranging from 0.027 μM to 50 μM.

The validation of the DT based models and self- consistency test were performed by 10-fold cross validation (CV) method. For classification models the sensitivity and specificity values were calculated The randomness of model was also determined by calculating Mathew’s correlation coefficient (MCC). The intercorrelation between estate contribution index (SssOE), augmented eccentric connectivity topochemical index (\(A\xi^{c}\)), molecular connectivity index (\(\chi^{4}\)) and connective eccentricity topochemical index (\(C^{4c}\)) was also investigated.

The following statistical measures were used to correlate biological activity and molecular descriptors for correlation models; n, number of molecules; k, number of descriptors in a model; df degree of freedom; $r^2$, coefficient of correlation; $q^2$, cross validated $r^2$; pred_ $r^2$, $r^2$ for external test set; pred_ $r^2$Se, coefficient of correlation of predicted dataset; Z score, Z score calculated by the randomization test; best _ran_ $r^2$; best _ran_ $q^2$, highest $q^2$ value in randomization test; $\alpha$, statistical significance parameter obtained by randomization test. Validation was done to study the internal stability and predictive ability of the correlation models. The significance of the models hence obtained was derived based on a calculated Z score (Gilbert, 1976; Golbraikh, A. Tropsha, 2003). The probability ($\alpha$) of significant of randomisation test is derived by comparing Z score value with Z score critical value.
The classification of 4-aminomethylene isoquinoline-1,3-dione analogues (Figure 8.1.) as inactive and active using a single tree, based on connective eccentricity topochemical index A5, augmented eccentric connectivity topochemical index A3, molecular connectivity index A1, alignment independent descriptor T_C_O_7 A19 and eccentric adjacency topochemical index A2 is shown in Figure 8.2.

![Decision Tree Diagram]

Figure 8.2. The decision tree for distinguishing active analogue (A) from inactive analogue (B); A5-connective eccentricity topochemical index, A3-augmented eccentric connectivity topochemical index, A1- molecular connectivity index, A19 - alignment independent descriptor T_C_O_7, A2- eccentric adjacency topochemical index (Gupta and Madan, 2013a)

The decision tree identified connective eccentricity topochemical index A5 as the most important index. The decision tree classified the analogues with an accuracy of 96%. The specificity and sensitivity of the training set was found to be of the order of 100% and 86.6% respectively (Table 8.3.). In 10 fold cross-validation, 69% of
Models for the prediction of CDK4 inhibitory activity

Aminomethylene isoquinoline-1, 3-diones analogues were correctly classified with regard to biological activity. The specificity and sensitivity of cross validated set was found to be 72.9% and 60% respectively (Table 8.3).

The RF classified aminomethylene isoquinoline-1,3-diones analogues as inactive and active with an accuracy of 96.4% with respect to CDK4 inhibitory activity. The out-of-bag (OOB) estimate of error was found to be only 8%. The specificity and sensitivity were of the order of 97.2% and 80% respectively and the value of MCC was found to be 0.7 as given in Table 8.3. High values of MCC simply indicate robustness of the proposed DT and RF based models for CDK4 inhibitory activities.

Table 8.3. Confusion matrix for CDK4 inhibitory activity using models based on decision tree and random forest

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Ranges</th>
<th>Number of compound predicted</th>
<th>Specifi city (%)</th>
<th>Sensitivity (%)</th>
<th>MCC</th>
<th>OOB (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision tree</td>
<td>Training set</td>
<td>Active</td>
<td>13</td>
<td>100</td>
<td>86.6</td>
<td>0.9</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inactive</td>
<td>0</td>
<td>100</td>
<td>86.6</td>
<td>0.9</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cross validated set</td>
<td>Active</td>
<td>9</td>
<td>72.9</td>
<td>60</td>
<td>0.3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inactive</td>
<td>10</td>
<td>72.9</td>
<td>60</td>
<td>0.3</td>
<td>-</td>
</tr>
<tr>
<td>Random forest</td>
<td>Active</td>
<td>12</td>
<td>97.2</td>
<td>80</td>
<td>0.7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inactive</td>
<td>1</td>
<td>97.2</td>
<td>80</td>
<td>0.7</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Using a single descriptor at a time, four independent MAA based models using E-state contribution index (SssOE), augmented eccentric connectivity topochemical index (\(A_{\xi_c}^c\)), molecular connectivity index (\(\chi^A\)) and connective eccentricity topochemical index (\(c^{\xi_c}\)) were developed (Table 8.1). The proposed models have been illustrated in Table 8.4.
Table 8.4. Proposed MAA based models for the prediction of CDK4 inhibitory activity

<table>
<thead>
<tr>
<th>Index</th>
<th>Nature of range</th>
<th>Index value</th>
<th>Total compounds in the range</th>
<th>Number of compounds predicted correctly</th>
<th>Overall accuracy of prediction</th>
<th>Average IC_{50} (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A24</td>
<td>Lower inactive</td>
<td>&lt; 24.868</td>
<td>17</td>
<td>17</td>
<td></td>
<td>5.17</td>
</tr>
<tr>
<td></td>
<td>Active</td>
<td>24.868 – 25.02</td>
<td>11</td>
<td>10</td>
<td>93.3</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Transitional</td>
<td>&gt; 25.02 – 25.1</td>
<td>8</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Upper inactive</td>
<td>&gt; 25.1 to ≥46.56</td>
<td>16</td>
<td>14</td>
<td>18.38</td>
<td>5.71</td>
</tr>
<tr>
<td>A10</td>
<td>Inactive</td>
<td>&lt; 19.28</td>
<td>21</td>
<td>21</td>
<td>90.6</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Transitional</td>
<td>19.28 to &lt; 20.46</td>
<td>20</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Active</td>
<td>20.46 to &lt; 22.04</td>
<td>11</td>
<td>8</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>A8</td>
<td>Lower inactive</td>
<td>&lt; 14.47</td>
<td>18</td>
<td>18</td>
<td></td>
<td>5.19</td>
</tr>
<tr>
<td></td>
<td>Lower transitional</td>
<td>14.47 to &lt; 14.89</td>
<td>8</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Active</td>
<td>14.895 to &lt; 15.26</td>
<td>8</td>
<td>7</td>
<td>94.4</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Upper inactive</td>
<td>15.26 to &lt; 16.07</td>
<td>9</td>
<td>8</td>
<td>7.33</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Upper transitional</td>
<td>≥ 16.07</td>
<td>9</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>A12</td>
<td>Inactive</td>
<td>&lt; 5.42</td>
<td>20</td>
<td>20</td>
<td></td>
<td>5.17</td>
</tr>
<tr>
<td></td>
<td>Transitional</td>
<td>5.42 to &lt; 5.64</td>
<td>25</td>
<td>NA</td>
<td>96.2</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Active</td>
<td>5.64 to &lt; 5.85</td>
<td>7</td>
<td>6</td>
<td>0.10</td>
<td>NA</td>
</tr>
</tbody>
</table>

NA: Not applicable

* Values in brackets are based upon correctly predicted analogues in the particular range

The overall accuracy of prediction varied from 90.6% for augmented eccentric connectivity topochemical index (A \( g^e \)) to 96.2% for connective eccentricity topochemical index (c^4c). Transitional ranges were observed in all the models indicating a gradual change in CDK4 inhibitory activity. The average IC_{50} (Table 8.4 and Figure 8.3) for active range in all the models varied from 0.15µM to 0.32 µM. The observation of extremely low average IC_{50} values indicates high potency of the active ranges in the proposed models. Consequently, these models offer vast potential for development of potent CDK4 inhibitors.
Figure 8.3. Average IC$_{50}$ (nM) values (based upon correctly predicted analogues) of 4-Aminomethylene isoquinoline-1, 3-dione derivatives for CDK4 inhibitory activity in various ranges of MAA based models (Gupta and Madan, 2013a).

Intercorrelation analysis (Table 8.5) revealed that E-state contribution index (SssOE), augmented eccentric connectivity topochemical index ($A^e_{C}$), molecular connectivity index ($A^c$) and connective eccentricity topochemical index ($C^c$) are not correlated with each other.

Table 8.5. Intercorrelation matrix for MDs used in MAA

<table>
<thead>
<tr>
<th></th>
<th>SssOE</th>
<th>$A^e_{C}$</th>
<th>$A^c$</th>
<th>$C^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SssOE</td>
<td>1</td>
<td>0.15</td>
<td>0.04</td>
<td>-0.2</td>
</tr>
<tr>
<td>$A^e_{C}$</td>
<td>1</td>
<td>0.46</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>$A^c$</td>
<td></td>
<td>1</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>$C^c$</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
After QSAR study by MLR using forward-backward stepwise variable selection method, the final equation developed and the statistical data observed are illustrated below.

\[
pIC_{50} = 0.1841 \times (T_{C\_O\_7}) - 0.4806 \times \text{(Rotatable Bond Count)} + 0.1247 \times \text{(SsOHE-index)} - 0.7727 \times (T_{2\_Cl\_7}) - 0.1639 \times \text{(SssOE-index)} + 7.4965 
\]

The QSAR model had a correlation coefficient \(r^2\) of 0.72, significant cross validated correlation coefficient \(q^2\) of 0.63, \(F\) test of 15.59, \(r^2\) for external test set \(\text{pred}_r^2\) 0.59, and degree of freedom 30. The model developed predicts 63\% of variance and is validated by an external set of compounds with a predictive correlation coefficient of 0.59. The model is validated by \(\alpha_{\text{ran}_r^2}= 0.00\), \(\alpha_{\text{ran}_q^2}= 0.001\), \(\alpha_{\text{ran}_\text{pred}_r^2}= 0.05\), \(\text{best}_\text{ran}_r^2= 0.31\), \(\text{best}_\text{ran}_q^2= 0.11\), \(Z\text{ score}_\text{ran}_r^2= 7.83\), \(Z\text{ score}_\text{ran}_q^2= 4.61\). The randomization test suggests that the developed model has a probability of less than 1\% and that the model is generated by chance. The predictability of model was evaluated by test set of compounds.

The reported and predicted \(pIC_{50}\) for training set along with residual values are presented in Table 8.6. The predictive ability of model evaluated using test set is presented in Table 8.7. The plot of reported vs predicted activity and contribution of descriptors for the CDK4 inhibitory activity is shown in Figure 8.4 and 8.5 respectively.

The major group of contributing descriptors involved subgroups like rotatable bond count, SsOHE-index, SssOE-index and alignment independent descriptors. These descriptors help in understanding the effect of substituent at different position of aminomethylene isoquinoline-1, 3(2\(H,4\(H\))-diones.
Table 8.6. Reported and predicted activity of 4-Aminomethylene isoquinoline-1, 3-dione derivatives used in training set with CDK4 inhibitory activity by MLR

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Index</th>
<th>R</th>
<th>IC₅₀ (μM)</th>
<th>pIC₅₀*</th>
<th>Residual</th>
<th>Reported</th>
<th>Predicted</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1a</td>
<td>H</td>
<td>4.1</td>
<td>5.38</td>
<td>6.1262</td>
<td>-0.739</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2a</td>
<td>7-Br</td>
<td>1.4</td>
<td>5.85</td>
<td>6.1262</td>
<td>-0.2723</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3a</td>
<td>6-Br</td>
<td>11</td>
<td>4.95</td>
<td>5.6047</td>
<td>-0.6461</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4a</td>
<td>6-NO₂</td>
<td>0.14</td>
<td>6.85</td>
<td>6.3819</td>
<td>0.472</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5a</td>
<td>6-I</td>
<td>11</td>
<td>4.95</td>
<td>5.1650</td>
<td>-0.2064</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6a</td>
<td>O=C-N(Me)₂</td>
<td>0.39</td>
<td>6.40</td>
<td>6.3819</td>
<td>0.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7a</td>
<td>N</td>
<td>0.33</td>
<td>6.48</td>
<td>6.0138</td>
<td>0.4677</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8a</td>
<td>O=N-CH₃</td>
<td>0.22</td>
<td>6.65</td>
<td>6.9340</td>
<td>-0.2764</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9a</td>
<td>N</td>
<td>0.13</td>
<td>6.88</td>
<td>6.2694</td>
<td>0.6167</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10a</td>
<td></td>
<td>14.3</td>
<td>4.84</td>
<td>5.5332</td>
<td>-0.6885</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11a</td>
<td>O</td>
<td>2.3</td>
<td>5.63</td>
<td>5.5332</td>
<td>0.1051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12a</td>
<td>O=O</td>
<td>10</td>
<td>5</td>
<td>5.2518</td>
<td>-0.2518</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1b</td>
<td>-</td>
<td>1.1</td>
<td>5.95</td>
<td>6.1262</td>
<td>-0.1676</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2b</td>
<td>-Br</td>
<td>2.5</td>
<td>5.60</td>
<td>5.3536</td>
<td>0.2485</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3b</td>
<td>-Cl</td>
<td>2</td>
<td>5.69</td>
<td>4.8577</td>
<td>0.8413</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4b</td>
<td>-OMe</td>
<td>15.1</td>
<td>4.82</td>
<td>5.6456</td>
<td>-0.8246</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>5b</td>
<td>-NO₂</td>
<td>0.3</td>
<td>6.52</td>
<td>6.1262</td>
<td>0.3967</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>6b</td>
<td>-I</td>
<td>21.2</td>
<td>4.6737</td>
<td>4.2038</td>
<td>0.4699</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>7b</td>
<td>O=C-N(Me)₂</td>
<td>0.14</td>
<td>6.8539</td>
<td>6.3819</td>
<td>0.472</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>8b</td>
<td>N</td>
<td>1</td>
<td>6</td>
<td>5.6456</td>
<td>0.3544</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>9b</td>
<td>O=N-CH₃</td>
<td>1.8</td>
<td>5.7447</td>
<td>5.5332</td>
<td>0.2115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>10b</td>
<td>N</td>
<td>0.32</td>
<td>6.4948</td>
<td>6.3819</td>
<td>0.1129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>11b</td>
<td>O</td>
<td>0.05</td>
<td>7.3010</td>
<td>6.3819</td>
<td>0.9191</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>12b</td>
<td>N</td>
<td>0.13</td>
<td>6.8861</td>
<td>6.9340</td>
<td>-0.0479</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. No.</td>
<td>Index</td>
<td>R</td>
<td>IC&lt;sub&gt;50&lt;/sub&gt; (μM)*</td>
<td>pIC&lt;sub&gt;50&lt;/sub&gt;&lt;sup&gt;#&lt;/sup&gt;</td>
<td>Residual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>------------</td>
<td>------------------------</td>
<td>-----------------------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Reported</td>
<td>Predicted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>13b</td>
<td>-OHN-</td>
<td>0.31</td>
<td>6.5086</td>
<td>6.3819</td>
<td>0.1267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>14b</td>
<td></td>
<td>19.4</td>
<td>4.7122</td>
<td>4.8365</td>
<td>-0.1243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>15b</td>
<td></td>
<td>0.027</td>
<td>7.5686</td>
<td>7.4952</td>
<td>0.0734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>16b</td>
<td></td>
<td>0.041</td>
<td>7.3872</td>
<td>7.4627</td>
<td>-0.0755</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>1c</td>
<td></td>
<td>50</td>
<td>4.3010</td>
<td>4.8204</td>
<td>-0.5194</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>2c</td>
<td></td>
<td>15.8</td>
<td>4.8013</td>
<td>5.9013</td>
<td>-1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>3c</td>
<td></td>
<td>2.7</td>
<td>5.5686</td>
<td>5.6456</td>
<td>-0.077</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>4c</td>
<td></td>
<td>3</td>
<td>5.5229</td>
<td>5.1650</td>
<td>0.3579</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>1d</td>
<td></td>
<td>0.11</td>
<td>6.9586</td>
<td>6.7500</td>
<td>0.2086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>2d</td>
<td></td>
<td>0.82</td>
<td>6.0862</td>
<td>6.7500</td>
<td>-0.6638</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>3d</td>
<td></td>
<td>0.08</td>
<td>7.0969</td>
<td>6.9340</td>
<td>0.1629</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>4d</td>
<td></td>
<td>0.25</td>
<td>6.602</td>
<td>6.5659</td>
<td>0.0361</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = Compound concentration in micro mole required to inhibit CDK4 activity by 50%.
# = -Log (IC<sub>50</sub> *10<sup>-6</sup>)
Models for the prediction of CDK4 inhibitory activity

Table 8.7. Reported and predicted activity of 4-Aminomethylene isoquinoline-1,3-dione derivatives used in test set for CDK4 inhibitory activity by MLR

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Index</th>
<th>R</th>
<th>IC₅₀ (µM)*</th>
<th>pIC₅₀ #</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reported</td>
<td>Predicted</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>13a</td>
<td></td>
<td>12.1</td>
<td>4.9172</td>
<td>6.1262</td>
</tr>
<tr>
<td>2.</td>
<td>14a</td>
<td></td>
<td>0.48</td>
<td>6.3188</td>
<td>6.1262</td>
</tr>
<tr>
<td>3.</td>
<td>15a</td>
<td></td>
<td>41</td>
<td>4.3872</td>
<td>4.2038</td>
</tr>
<tr>
<td>4.</td>
<td>16a</td>
<td></td>
<td>1.62</td>
<td>5.7905</td>
<td>6.3819</td>
</tr>
<tr>
<td>5.</td>
<td>18a</td>
<td></td>
<td>0.1</td>
<td>7</td>
<td>5.9013</td>
</tr>
<tr>
<td>6.</td>
<td>18b</td>
<td></td>
<td>29</td>
<td>4.5376</td>
<td>5.5510</td>
</tr>
<tr>
<td>7.</td>
<td>19b</td>
<td></td>
<td>3.3</td>
<td>5.4815</td>
<td>6.1262</td>
</tr>
<tr>
<td>8.</td>
<td>20b</td>
<td></td>
<td>3.5</td>
<td>5.4559</td>
<td>5.6047</td>
</tr>
<tr>
<td>9.</td>
<td>21b</td>
<td></td>
<td>0.92</td>
<td>6.0362</td>
<td>6.3819</td>
</tr>
<tr>
<td>10.</td>
<td>22b</td>
<td></td>
<td>34.7</td>
<td>4.4597</td>
<td>5.9013</td>
</tr>
<tr>
<td>11.</td>
<td>23b</td>
<td></td>
<td>0.037</td>
<td>7.4318</td>
<td>6.9340</td>
</tr>
<tr>
<td>12.</td>
<td>24b</td>
<td></td>
<td>27.8</td>
<td>4.5560</td>
<td>4.9132</td>
</tr>
<tr>
<td>13.</td>
<td>25b</td>
<td></td>
<td>0.32</td>
<td>6.4948</td>
<td>6.1262</td>
</tr>
<tr>
<td>14.</td>
<td>5c</td>
<td></td>
<td>2.8</td>
<td>5.5528</td>
<td>5.1650</td>
</tr>
<tr>
<td>15.</td>
<td>5d</td>
<td></td>
<td>0.13</td>
<td>6.8862</td>
<td>4.9292</td>
</tr>
<tr>
<td>16.</td>
<td>6d</td>
<td></td>
<td>1.25</td>
<td>5.9031</td>
<td>6.75</td>
</tr>
</tbody>
</table>

* = Compound concentration in micro mole required to inhibit CDK4 activity by 50%.
# = -Log (IC₅₀ *10⁻⁶)
The direct relationship of descriptor T_C_O_7 (27.23%) suggested that presence of oxo group at position one in the basic ring 4-aminomethylene-4-H isoquinoline-1,3-dione should necessarily be separated from the substituent at position six by seven bonds.

The presence of T_2_Cl_7 (having negative MLR coefficient – 16.43%) in the model revealed that the presence of halides at position seven i.e. chloro at position seven separated by seven bonds from doubly bonded oxo at position three have negative effect on the activity.
The directly related descriptor SsOHE-index (16.22%) indicated that the presence of hydroxyl group in six phenyl substituted basic ring have positive effect on the activity. SssOE-index (-14.6%) is an estate contribution descriptor which represent electrotopological estate indices for the number of oxygen group connect with two single bonds. This term is negatively correlated and indicated that compound with higher SssOE-index values show less activity and vice-versa.

The presence of descriptor rotatable bond count (having negative MLR coefficient -25.53 %) signifies that the unsaturated bonds, and single bonds connected to hydrogen or terminal atoms are favorable for the biological activity.
After QSAR study by PLS using forward-backward stepwise variable selection method, the final equation developed and the statistical data observed are illustrated below.

$$pIC_{50} = 0.1827 \ (T_{C\ O\ 7}) - 0.4797 \ (\text{Rotatable bond count}) + 0.1253 \ (\text{SsOHE-index}) - 0.7850 \ (T_{2\ Cl\ 7}) - 0.1645 \ (\text{SssOE-index}) + 7.5007$$

The QSAR model had a correlation coefficient ($r^2$) of 0.72, significant cross validated correlation coefficient ($q^2$) of 0.6, F test of 27.72, $r^2$ for external test set (pred $r^2$) 0.53, and degree of freedom 32. The model was validated by $\alpha_{\text{ran} \ r^2}= 0.00$, $\alpha_{\text{ran} \ q^2}= 0.00$, $\alpha_{\text{ran} \ pred \ r^2}= 0.01$, best $\alpha_{\text{ran} \ r^2}= 0.38$, best $\alpha_{\text{ran} \ q^2}= 0.21$, Z score $\alpha_{\text{ran} \ r^2}= 7.41$, Z score $\alpha_{\text{ran} \ q^2}= 4.57$.

The plot of reported vs predicted activity and contribution of descriptors for the CDK4 inhibitory activity is shown in Figure 8.6. The major groups of descriptors involved in developing the equation by PLSR are subgroups like rotatable bond count, SsOHE-index, SssOE-index and alignment independent descriptors. The descriptors are common between MLR and PLSR. These only differ from each other in their percentage of contribution.

![Figure 8.6](image_url)

**Figure 8.6.** (A) Graph of Reported vs. Predicted activities for training and test set molecules by Partial Least Square (PLS) Regression model. A) Training set (solid dots) (B) Test Set (hollow dots). (B) Percentage contribution of each descriptor in developed PLS model explaining variation in the activity (Gupta and Madan, 2013a)
Models for the prediction of CDK4 inhibitory activity

After QSAR study by PCR using forward-backward stepwise variable selection method, the final equation developed and the statistical data observed are illustrated below.

\[ \text{pIC}_{50} = 7.2205 \times (\text{chi5chain}) + 0.1097 \times (T\_C\_O\_7) - 0.3324 \times (T\_2\_N\_4) + 0.1682 \times (\text{SsOHE-index}) - 0.2061 \times (T\_2\_O\_3) + 9.2238 \]

The QSAR model had a correlation coefficient \((r^2)\) of 0.72, significant cross validated correlation coefficient \((q^2)\) of 0.6, F test of 20.24 , \(r^2\) for external test set \((\text{pred}_r^2)\) 0.52 , and degree of freedom 30. The model is validated by \(\alpha_{\text{ran}_r^2}=0.00\), \(\alpha_{\text{ran}_q^2}=0.00\), \(\alpha_{\text{ran}_\text{pred}_r^2}=0.001\), best\_ran\_r^2=0.34, best\_ran\_q^2=0.2.

The plot of reported vs predicted activity and contribution of descriptors for the CDK4 inhibitory activity is shown in Figure 8.7.

![Graph](image1.png)

![Bar Chart](image2.png)

**Figure 8.7.** (A) Graph of Reported vs. Predicted activities for training and test set molecules by Principal Component (PCR) Regression model. A) Training set (solid dots) and Test Set (hollow dots). (B) Plot of percentage contribution of each descriptor in developed PCR model explaining variation in the activity (Gupta and Madan, 2013a)
The major groups of descriptors involved in developing the equation by PCR are subgroups like chi5chain, SsOHE-index and alignment independent descriptors. The QSAR model by PCR reveals that the descriptors SsOHE-index and T_C_O_7 are common in MLR and PCR. These only differ from each other in their percentage of contribution. The other contributing descriptors are the chi5chain, T_2_N_4 and T_2_O_3.

The chi5chain (28.25%) is directly proportional to the biological activity. The descriptor T_2_N_4 (-19.53) is negatively correlated with activity shows that increasing the distance between 4-amino phenylmethylene from oxo by increasing number of carbon atoms in the basic ring have negative effect on the activity. The descriptor T_2_O_3 (-12.63) which is also negatively correlated with activity shows that the increase and decrease of distance between two oxo group have negative effect on the activity. Combined approaches using molecular properties and well selected MDs are likely not only to produce superior correlations but are expected to do so in a most efficient way. Structure-activity studies are highly complex and various methodologies, even if addressing limited aspects of the QSAR problem, ought to be exhaustively explored and amalgamated if possible. High accuracy of prediction offers proposed models a vast potential for providing lead structures for the development of potent therapeutic agents for CDK4 inhibition.