

## 化学物質の発ガン性予測 SVM モデルにおける 電子的記述子の有効性の検討

鈴木孝弘<sup>a</sup>・松本高利<sup>b</sup>・田辺和俊<sup>c</sup>

### Verification of Effectiveness of Electronic Descriptors in SVM Models for Predicting Carcinogenicities of Chemicals

Takahiro SUZUKI<sup>a</sup>, Takatoshi MATSUMOTO<sup>b</sup>,  
Kazutoshi TANABE<sup>c</sup>

#### Abstract

It has been demonstrated that our parallel model combining support vector machine (SVM) models constructed for twenty substructure groups predicts the carcinogenicities of a wide variety of chemicals with a satisfactory overall accuracy of approximately 80%. Besides, the methods of variable selection in SVM modeling were tested in order to improve the performance of this model by raising the accuracy for N-nitroso-, nitroso- and nitroaromatic groups (89 chemicals) which showed the lowest accuracy (70.8%) among twenty molecular groups. It was found that the sensitivity analysis method improves the accuracy of the group from 70.8 % to 77.5 %. Unfortunately, this level is not satisfactory as the performance of a carcinogenicity prediction tool alternative to animal tests. In this study, the effect of the employment of intrinsic electronic descriptors such as HOMO and LUMO energies and electronic density on the prediction performance of the above model was examined to improve the predictive power of the SVM model, especially for N-nitroso-, nitroso- and nitroaromatic groups.

<sup>a</sup> 東洋大学自然科学研究室 : 〒 112-8606 東京都文京区白山 5-28-20

Natural Science Lab., Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606 JAPAN

<sup>b</sup> 東北大学多元物質科学研究所 : 〒 980-8577 宮城県仙台市青葉区片平 2-1-1

Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577 JAPAN

<sup>c</sup> 産業技術総合研究所ヒューマンライフテクノロジー研究部門 : 〒 305-8568 茨城県つくば市梅園 1-1-1 つくば中央第2

Human Technology Research Institute, National Institute of Advanced Industrial Science and Technology, AIST Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568 JAPAN

**Keywords:** carcinogenicity database; carcinogenicity prediction; molecular descriptors; quantitative structure-activity relationship (QSAR); support vector machine (SVM)

## 1. はじめに

化学物質の発ガン性を評価するための動物試験の代替として、著者らは定量的構造活性相関 (QSAR) により多種多様な構造の化学物質の発ガン性を構造から予測するモデルを構築し、同族体群ごとに構築したサポートベクターマシン (SVM) を並列に組み合わせるモデルが約900種の化学物質の発ガン性を80%の精度で予測できることを見出した (Tanabe et al, 2010; Tanabe et al, 2011a). さらに、このモデルの予測性能を向上させるために、同族体群の中で正解率が他の群に比べて最も低い芳香族N-ニトロソ・ニトロソ・ニトロ群 (89物質) についてSVMモデルにおける記述子選択の方法を検討した. 相関係数法, F-score法, 感度分析法の3手法により選択した記述子を用いて正解率を調べたところ、感度分析法を用いることによって、この群の正解率が前報の70.8%から77.5%に向上することを見出した (Tanabe et al, 2011b; Suzuki et al, 2012). しかし、この77.5%という数値は、動物実験の代替としての発ガン性予測手法の正解率としては未だ満足できる数値ではない.

この芳香族N-ニトロソ・ニトロソ・ニトロ群については、芳香環の置換基の位置や種類の違いによりN-ニトロソ・ニトロソ・ニトロ基の電子密度や分子全体の電子軌道エネルギー・dipole moment等が変化することが予測され、そのような電子状態の違いが化合物の発ガン性の違いを引き起こす可能性が考えられる (Tanabe et al, 2010). その他の化合物についての発ガン性と電子状態の関係については、1939年、Schmidtは発ガン性と芳香族化合物の電子構造に関するという仮説を発表し、1945年、Pullmanは縮合多環芳香族化合物のK領域の局在化エネルギーと発ガン性の関係を報告している。それ以来、電子状態を表す記述子を取り込んだ発ガン性予測モデルについては、化学構造が類似した化合物群については非常に多くの研究がある (Vracko, 1997; Gini et al, 1999a, 1999b; Vendrame et al, 1999; Greene, 2002; Contrera et al, 2003; Hemmateenejad et al, 2004; Benigni, 2005; Bohari et al, 2011; Fioravanzo et al, 2012; Fjodorova et al, 2012, 等).

そこで、本報では、芳香族N-ニトロソ・ニトロソ・ニトロ群の89分子について分子軌道法により電子状態を計算し、上記の予測モデルに幾つかの電子状態を表す記述子を追加して解析を行い、発ガン性の予測率が向上するか否かを検討し、電子的記述子の発ガン性予測での有効性を検証した.

## 2. 方法

### 2.1 解析対象の化学物質

本報で解析した化学物質は前報 (Tanabe et al, 2010; Tanabe et al, 2011a) で用いたのと同じ89種の芳香族N-ニトロソ・ニトロソ・ニトロ化合物である。それらの構造式, 化学名, 化学式, 発ガン性ランクをFig. 1 a~cに示す。発ガン性ランクは動物実験に基づく発ガン性の陽性/陰性とその信頼性を表し, A~Cが陽性, DとEが陰性であり, Aが最も信頼性が高く, BとEは中程度, CとDは信頼性が最も低いことを表す(詳細については前報 (Tanabe et al, 2010) を参照)。このFig. 1 a~cを見ると, 化学式が同一で置換基の位置が異なる異性体の組が幾つか存在し, さらにその中には発ガン性ランクの異なる化学物質が存在する組があることがわかる。

### 2.2 分子の記述子

以上の89物質について, 前報 (Tanabe et al, 2010; Tanabe et al, 2011a) で用いた1,504種のDragon記述子に, 今回新たに作成した電子状態記述子を追加してQSAR(定量的構造活性相関)解析を行った。新規に追加した記述子はGaussian98プログラムを用い, 基底関数4-31G, RHFにより電子状態を計算し, Table 1 a,bに示す10種の特性量を用いた。この表には, 上記の異性体の組についての電子状態記述子の数値を示したが, それら異性体間で値が大きく異なる記述子がある一方, 値がほとんど変わらない記述子もあることが分かる。これらの記述子を追加して解析することで発ガン性の予測正解率が前報 (Tanabe et al, 2011b; Suzuki et al, 2012) より向上するか否かを検討した。

ただし, これらの記述子は物質数に比べて明らかに過多である。そのため, 予測に有効かつ不可欠な説明変数をスクリーニングする必要がある。本稿では, 前報 (Tanabe et al, 2011b; Suzuki et al, 2012) で提案した感度分析法を用いて, 記述子の選定を試みた。

### 2.3 モデルの作成とその最適化

前報と同じSVMのソフトウェアLIBSVM ver.2.89 (Chang et al, 2009a) を用いて, 発ガン性データと記述子データとの相関解析モデルを作成した。発ガン性データはTable 2のように信頼性に対応させた重みを設定し, LIBSVMの重み付きデータに対する2群分類機能 (Chang et al, 2009b) を用いた。

前報と同じDual Cross-Validation Testにより, SVMの学習・テストと最適化を同時に次の手順で実施した。

- ① 解析対象の化学物質を10群に分割する,
- ② その内の9群を学習用とし, この群についてLeave-One-Outを用いてパラメータ $g(\gamma)$ と $c(\text{cost})$ および記述子数を最適化する,
- ③ その最適モデルを用いて, テスト用物質の発ガン性を予測する,

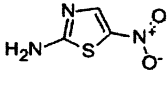
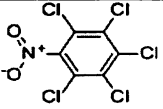
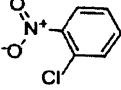
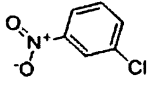
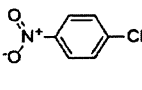
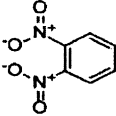
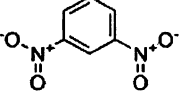
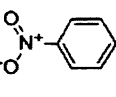
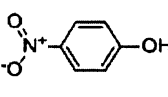
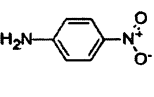
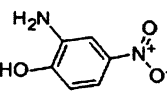
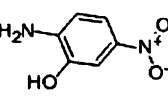
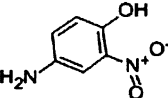
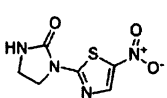
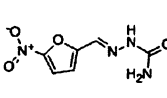
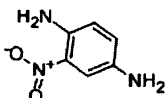
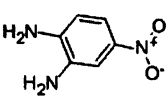
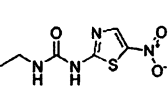
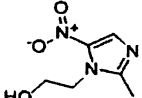
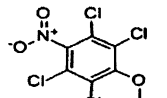
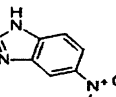
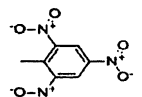
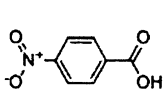
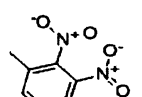
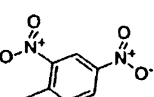
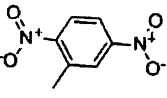
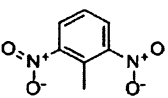
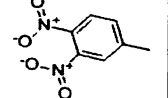
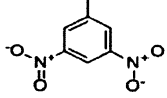
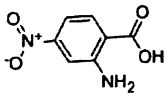
 2-Amino-5-nitrothiazole $C_3H_3N_3O_2S$ D	 Pentachloronitrobenzene $C_6Cl_5NO_2$ D	 1-Chloro-2-nitrobenzene $C_6H_4ClNO_2$ D	 1-Chloro-3-nitrobenzene $C_6H_4ClNO_2$ D	 1-Chloro-4-nitrobenzene $C_6H_4ClNO_2$ C
 o-Dinitrobenzene $C_6H_4N_2O_4$ D	 m-Dinitrobenzene $C_6H_4N_2O_4$ D	 Nitrobenzene $C_6H_5NO_2$ B	 p-Nitrophenol $C_6H_5NO_3$ E	 p-Nitroaniline $C_6H_6N_2O_2$ D
 2-Amino-4-nitrophenol $C_6H_6N_2O_3$ D	 2-Amino-5-nitrophenol $C_6H_6N_2O_3$ D	 4-Amino-2-nitrophenol $C_6H_6N_2O_3$ D	 Nitridazole $C_6H_6N_4O_3S$ B	 Furacilin $C_6H_6N_4O_4$ D
 2-Nitro-p-phenylenediamine $C_6H_7N_3O_2$ D	 4-Nitro-o-phenylenediamine $C_6H_7N_3O_2$ D	 1-Ethyl-3-(5-nitro-2-thiazolyl)urea $C_6H_8N_4O_3S$ D	 Metronidazole $C_6H_9N_3O_3$ B	 2,3,5,6-Tetrachloro-4-nitroanisole $C_7H_3Cl_4NO_3$ E
 6-Nitrobenzimidazole $C_7H_5N_3O_2$ E	 2,4,6-Trinitrotoluene $C_7H_5N_3O_6$ C	 p-Nitrobenzoic acid $C_7H_5NO_4$ E	 2,3-Dinitrotoluene $C_7H_6N_2O_4$ B	 2,4-Dinitrotoluene $C_7H_6N_2O_4$ B
 2,5-Dinitrotoluene $C_7H_6N_2O_4$ B	 2,6-Dinitrotoluene $C_7H_6N_2O_4$ B	 3,4-Dinitrotoluene $C_7H_6N_2O_4$ B	 3,5-Dinitrotoluene $C_7H_6N_2O_4$ B	 4-Nitroanthranilic acid $C_7H_6N_2O_4$ E

Fig. 1a. Chemical structures, names, formulae, and carcinogenicity ranks of 89 chemicals treated in this study

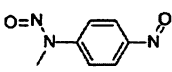
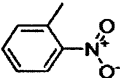
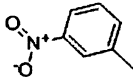
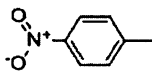
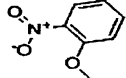
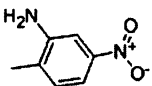
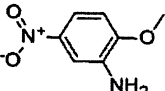
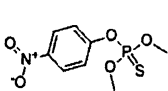
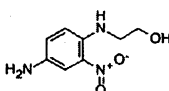
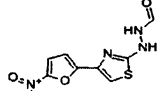
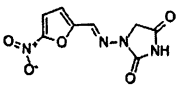

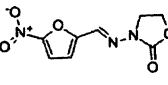
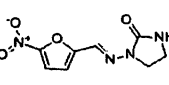
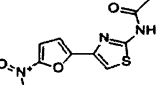
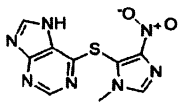
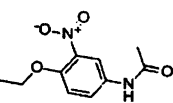
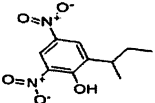
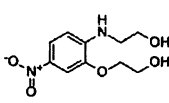
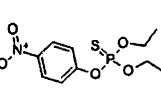
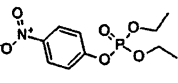
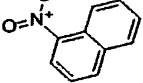
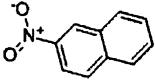
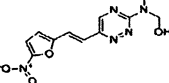
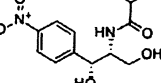
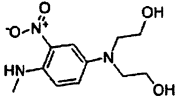
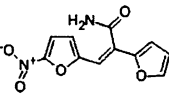
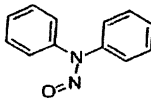
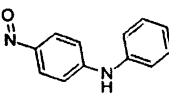
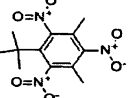
 <p>N-Methyl-N,4-dinitrosoaniline C7H7N3O2 D</p>	 <p>o-Nitrotoluene C7H7NO2 B</p>	 <p>m-Nitrotoluene C7H7NO2 D</p>	 <p>p-Nitrotoluene C7H7NO2 D</p>	 <p>o-Nitroanisole C7H7NO3 B</p>
 <p>5-Nitro-o-toluidine C7H8N2O2 C</p>	 <p>5-Nitro-o-anisidine C7H8N2O3 D</p>	 <p>Methylparathion C8H10NO5PS D</p>	 <p>HC Red No.3 C8H11N3O3 D</p>	 <p>Nifurthiazole C8H6N4O4S B</p>
 <p>Nitrofurantoin C8H6N4O5 D</p>	 <p>Dinitolmide C8H7N3O5 D</p>	 <p>Furazolidone C8H7N3O5 D</p>	 <p>Nifuradene C8H8N4O4 B</p>	 <p>Furothiazole C9H7N3O4S B</p>
 <p>Azathioprine C9H7N7O2S A</p>	 <p>3-Nitro-p-acetophenide C10H12N2O4 E</p>	 <p>6-sec-Butyl-2,4-dinitrophenol C10H12N2O5 D</p>	 <p>N,O-Di(2-hydroxyethyl)-2-amino-5-nitrophenol C10H14N2O5 D</p>	 <p>Parathion C10H14NO5PS C</p>
 <p>Diethyl p-nitrophenyl phosphorate C10H14NO6P D</p>	 <p>1-Nitronaphthalene C10H7NO2 D</p>	 <p>2-Nitronaphthalene C10H7NO2 B</p>	 <p>Bis(hydroxymethyl)-furatrizine C11H11N5O5 D</p>	 <p>Chloramphenicol C11H12Cl2N2O5 B</p>
 <p>HC Blue No.1 C11H17N3O4 B</p>	 <p>AF-2 C11H8N2O5 B</p>	 <p>N-Nitrosodiphenylamine C12H10N2O C</p>	 <p>p-Nitrosodiphenylamine C12H10N2O D</p>	 <p>Musk xylene C12H15N3O6 C</p>

Fig. 1b. Chemical structures, names, formulae, and carcinogenicity ranks of 89 chemicals treated in this study

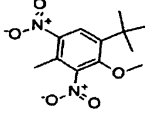
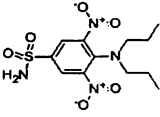
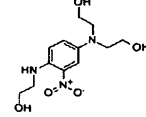
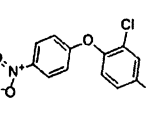
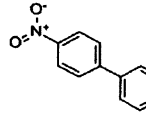
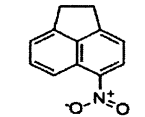
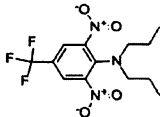
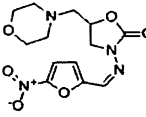
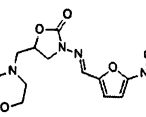
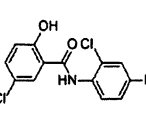
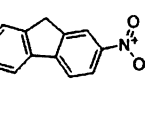
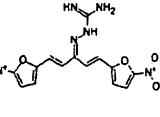
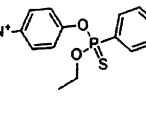
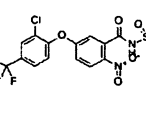
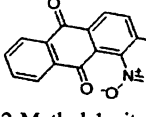
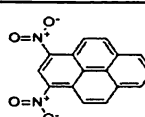
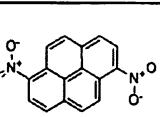
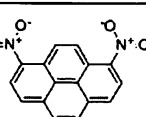
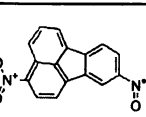
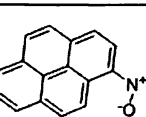
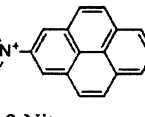
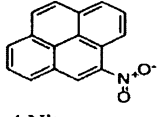
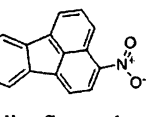
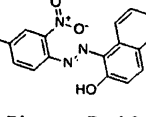
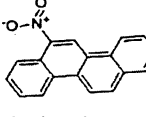
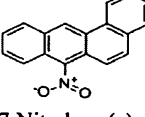
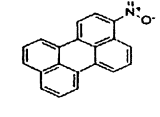
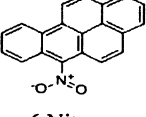
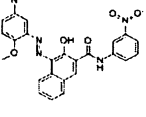
 <p><b>Musk ambrette</b> C<sub>12</sub>H<sub>16</sub>N<sub>2</sub>O<sub>5</sub> D</p>	 <p><b>Oryzalin</b> C<sub>12</sub>H<sub>18</sub>N<sub>4</sub>O<sub>6</sub>S C</p>	 <p><b>HC Blue 2</b> C<sub>12</sub>H<sub>19</sub>N<sub>3</sub>O<sub>5</sub> D</p>	 <p><b>Nitrofen</b> C<sub>12</sub>H<sub>7</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>3</sub> B</p>	 <p><b>4-Nitrobiphenyl</b> C<sub>12</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub> B</p>
 <p><b>5-Nitroacenaphthene</b> C<sub>12</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub> B</p>	 <p><b>Trifluralin</b> C<sub>13</sub>H<sub>16</sub>F<sub>3</sub>N<sub>3</sub>O<sub>4</sub> C</p>	 <p><b>Levofuraltadone</b> C<sub>13</sub>H<sub>16</sub>N<sub>4</sub>O<sub>6</sub> B</p>	 <p><b>Furaltadone</b> C<sub>13</sub>H<sub>16</sub>N<sub>4</sub>O<sub>6</sub> B</p>	 <p><b>Clonitralid</b> C<sub>13</sub>H<sub>7</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>4</sub> E</p>
 <p><b>2-Nitrofluorene</b> C<sub>13</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub> B</p>	 <p><b>Nitrovin</b> C<sub>14</sub>H<sub>12</sub>N<sub>6</sub>O<sub>6</sub> D</p>	 <p><b>EPN</b> C<sub>14</sub>H<sub>14</sub>N<sub>4</sub>O<sub>4</sub>PS D</p>	 <p><b>Fomesafen</b> C<sub>15</sub>H<sub>10</sub>ClF<sub>3</sub>N<sub>2</sub>O<sub>6</sub>S C</p>	 <p><b>2-Methyl-1-nitroanthraquinone</b> C<sub>15</sub>H<sub>9</sub>N<sub>2</sub>O<sub>4</sub> B</p>
 <p><b>1,3-Dinitropyrene</b> C<sub>16</sub>H<sub>8</sub>N<sub>2</sub>O<sub>4</sub> D</p>	 <p><b>1,6-Dinitropyrene</b> C<sub>16</sub>H<sub>8</sub>N<sub>2</sub>O<sub>4</sub> B</p>	 <p><b>1,8-Dinitropyrene</b> C<sub>16</sub>H<sub>8</sub>N<sub>2</sub>O<sub>4</sub> B</p>	 <p><b>3,9-Dinitrofluoranthene</b> C<sub>16</sub>H<sub>8</sub>N<sub>2</sub>O<sub>4</sub> B</p>	 <p><b>1-Nitropyrene</b> C<sub>16</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub> B</p>
 <p><b>2-Nitropyrene</b> C<sub>16</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub> D</p>	 <p><b>4-Nitropyrene</b> C<sub>16</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub> B</p>	 <p><b>3-Nitrofluoranthene</b> C<sub>16</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub> D</p>	 <p><b>Pigment Red 3</b> C<sub>17</sub>H<sub>13</sub>N<sub>3</sub>O<sub>3</sub> D</p>	 <p><b>6-Nitrochrysene</b> C<sub>18</sub>H<sub>11</sub>N<sub>2</sub>O<sub>2</sub> B</p>
 <p><b>7-Nitrobenzo(a)anthracene</b> C<sub>18</sub>H<sub>11</sub>N<sub>2</sub>O<sub>2</sub> D</p>	 <p><b>3-Nitroperylene</b> C<sub>20</sub>H<sub>11</sub>N<sub>2</sub>O<sub>2</sub> D</p>	 <p><b>6-Nitrobenzo(a)pyrene</b> C<sub>20</sub>H<sub>11</sub>N<sub>2</sub>O<sub>2</sub> D</p>	 <p><b>C.I. Pigment Red 23</b> C<sub>24</sub>H<sub>17</sub>N<sub>5</sub>O<sub>7</sub> E</p>	

Fig. 1c. Chemical structures, names, formulae, and carcinogenicity ranks of 89 chemicals treated in this study

**Table 1a.** Electronic descriptors and their values in several isomeric pairs

Chemical Name	Mol. Form.	Rank	HOMO	LUMO	Gap	DM	ED(N)
1-Chloro-2-nitrobenzene	C6H4ClNO2	D	-0.3773	0.0383	0.4156	6.5183	6.6952
1-Chloro-3-nitrobenzene	C6H4ClNO2	D	-0.3801	0.0247	0.4047	4.8982	6.6990
1-Chloro-4-nitrobenzene	C6H4ClNO2	C	-0.3877	0.0253	0.4130	3.3344	6.7057
<i>o</i> -Dinitrobenzene	C6H4N2O4	D	-0.4131	0.0098	0.4229	8.7299	6.7040
<i>m</i> -Dinitrobenzene	C6H4N2O4	D	-0.4148	0.0043	0.4191	5.4394	6.7144
2-Amino-4-nitrophenol	C6H6N2O3	D	-0.3088	0.0473	0.3561	6.5023	6.7026
2-Amino-5-nitrophenol	C6H6N2O3	D	-0.3149	0.0499	0.3648	6.6189	6.7070
4-Amino-2-nitrophenol	C6H6N2O3	D	-0.3036	0.0587	0.3624	8.0242	6.6817
2-Nitro- <i>p</i> -phenylenediamine	C6H7N3O2	D	-0.2766	0.0480	0.3246	6.4996	6.7391
4-Nitro- <i>o</i> -phenylenediamine	C6H7N3O2	D	-0.3149	0.0574	0.3723	7.5860	6.7103
2,3-Dinitrotoluene	C7H6N2O4	B	-0.3955	0.0154	0.4108	8.7615	6.7628
2,4-Dinitrotoluene	C7H6N2O4	B	-0.4030	0.0113	0.4143	6.1857	6.7302
2,5-Dinitrotoluene	C7H6N2O4	B	-0.3936	-0.0090	0.3846	1.0029	6.7271
2,6-Dinitrotoluene	C7H6N2O4	B	-0.3990	0.0187	0.4176	3.8910	6.7127
3,4-Dinitrotoluene	C7H6N2O4	B	-0.3997	0.0147	0.4144	9.3465	6.7149
3,5-Dinitrotoluene	C7H6N2O4	B	-0.3971	0.0070	0.4041	6.0471	6.7189
<i>o</i> -Nitrotoluene	C7H7NO2	B	-0.3603	0.0446	0.4048	5.5048	6.7174
<i>m</i> -Nitrotoluene	C7H7NO2	D	-0.3603	0.0405	0.4008	6.0759	6.7013
<i>p</i> -Nitrotoluene	C7H7NO2	D	-0.3680	0.0426	0.4106	6.4104	6.7024
1-Nitronaphthalene	C10H7NO2	D	-0.3193	0.0349	0.3542	5.8435	6.7208
2-Nitronaphthalene	C10H7NO2	B	-0.3197	0.0333	0.3530	6.4855	6.7083
<i>N</i> -Nitrosodiphenylamine	C12H10N2O	C	-0.3173	0.1076	0.4249	4.2207	6.5635
<i>p</i> -Nitrosodiphenylamine	C12H10N2O	D	-0.2992	0.0674	0.3666	6.7223	6.7885
1,3-Dinitropyrene	C16H8N2O4	D	-0.3113	-0.0093	0.3020	9.6184	6.7387
1,6-Dinitropyrene	C16H8N2O4	B	-0.3139	-0.0093	0.3046	0.0009	6.7337
1,8-Dinitropyrene	C16H8N2O4	B	-0.3137	-0.0083	0.3055	9.0455	6.7372
1-Nitropyrene	C16H9NO2	B	-0.2863	0.0220	0.3083	6.9010	6.7308
2-Nitropyrene	C16H9NO2	D	-0.2822	0.0345	0.3167	6.9647	6.7147
4-Nitropyrene	C16H9NO2	B	-0.2821	0.0288	0.3109	6.2655	6.7300

Table 1b. Electronic descriptors and their values in several isomeric pairs

Chemical Name	Mol. Form.	Rank	ED(O1)	ED(O2)	MA	AA	SEV
1-Chloro-2-nitrobenzene	C6H4ClNO2	D	8.5510	8.5296	145.58	313.46	131.41
1-Chloro-3-nitrobenzene	C6H4ClNO2	D	8.5600	8.5575	148.15	320.35	131.50
1-Chloro-4-nitrobenzene	C6H4ClNO2	C	8.5599	8.5599	148.11	320.25	131.48
<i>o</i> -Dinitrobenzene	C6H4N2O4	D	8.5318	8.5311	154.51	327.76	141.77
<i>m</i> -Dinitrobenzene	C6H4N2O4	D	8.5597	8.5500	157.06	334.23	140.68
2-Amino-4-nitrophenol	C6H6N2O3	D	8.5655	8.5652	149.45	321.49	132.83
2-Amino-5-nitrophenol	C6H6N2O3	D	8.5773	8.5698	149.41	321.45	132.77
4-Amino-2-nitrophenol	C6H6N2O3	D	8.5641	8.5049	149.67	322.17	133.28
2-Nitro- <i>p</i> -phenylenediamine	C6H7N3O2	D	8.6181	8.5445	150.09	321.57	134.15
4-Nitro- <i>o</i> -phenylenediamine	C6H7N3O2	D	8.5779	8.5755	152.02	326.59	135.03
2,3-Dinitrotoluene	C7H6N2O4	B	8.5609	8.5429	167.79	346.48	158.40
2,4-Dinitrotoluene	C7H6N2O4	B	8.5767	8.5419	170.49	354.11	156.42
2,5-Dinitrotoluene	C7H6N2O4	B	8.5606	8.5393	170.66	354.57	156.63
2,6-Dinitrotoluene	C7H6N2O4	B	8.5502	8.5410	168.08	348.21	156.52
3,4-Dinitrotoluene	C7H6N2O4	B	8.5357	8.5350	171.23	356.62	158.07
3,5-Dinitrotoluene	C7H6N2O4	B	8.5602	8.5494	173.77	363.05	157.01
<i>o</i> -Nitrotoluene	C7H7NO2	B	8.5790	8.5550	142.41	307.35	128.22
<i>m</i> -Nitrotoluene	C7H7NO2	D	8.5632	8.5627	145.73	316.37	128.83
<i>p</i> -Nitrotoluene	C7H7NO2	D	8.5657	8.5654	145.72	316.31	128.81
1-Nitronaphthalene	C10H7NO2	D	8.5627	8.5458	168.11	346.86	158.21
2-Nitronaphthalene	C10H7NO2	B	8.5655	8.5619	170.78	353.75	158.64
<i>N</i> -Nitrosodiphenylamine	C12H10N2O	C	8.4132	8.4132	199.95	401.65	192.70
<i>p</i> -Nitrosodiphenylamine	C12H10N2O	D	8.4027	8.4027	203.27	411.16	190.59
1,3-Dinitropyrene	C16H8N2O4	D	8.5652	8.5346	229.64	428.40	196.59
1,6-Dinitropyrene	C16H8N2O4	B	8.5635	8.5362	245.86	463.37	249.58
1,8-Dinitropyrene	C16H8N2O4	B	8.5581	8.5400	245.92	463.46	249.60
1-Nitropyrene	C16H9NO2	B	8.5679	8.5475	220.47	423.45	221.73
2-Nitropyrene	C16H9NO2	D	8.5624	8.5624	223.27	430.65	222.24
4-Nitropyrene	C16H9NO2	B	8.5599	8.5461	220.60	423.81	221.78

HOMO : HOMO energy, LUMO : LUMO energy, Gap : HOMO-LUMO energy gap, DM : dipole moment, ED(N) : electron density on N atom of N-nitroso, nitroso or nitro group, ED(O1) : electron density on O atom of N-nitroso or nitroso group, or electron density on higher O atom of nitro group, ED(O2) : electron density on O atom of N-nitroso or nitroso group, or electron density on lower O atom of nitro group, MA : molecular area, AA : accessible area, SEV : solvent exclusive area.



**Table 2.** Target and weight values of carcinogenicity ranks in LIBSVM

Carcinogenicity	Rank	Number of chemicals	Target value	Weight value
Positive	A	1	1.0	1.00
	B	31	1.0	0.50
	C	9	1.0	0.25
	Total	41		
Negative	D	40	-1.0	0.25
	E	8	-1.0	0.50
	F	0	-1.0	1.00
	Total	48		
Total		89		

**Table 3.** Results of performances of three models

Model	Ref	Number of optimized descriptors	Overall accuracy
I: Parallel SVM model	[Tanabe et al, 2010, 2011a]	75	0.708
II: I + Sensitivity analysis	[Tanabe et al, 2011b, 2012]	35	0.775
III: II + Electronic descriptors	This work	41	0.775

**Table 4.** Orders of contribution of ten electronic descriptors to carcinogenicity

Descriptor	Order	Descriptor	Order	Descriptor	Order
HOMO	184	ED(O2)	below 300	MA	below 300
ED(O1)	215	DM	below 300	AA	below 300
LUMO	259	SEV	below 300		
ED(N)	276	Gap	below 300		

- ④ 以上の手順を学習用とテスト用物質を入れ替えながら10回繰り返し、全ての物質について発ガン性を予測する、  
モデルの性能評価には、次式の総合正解率 (Overall Accuracy; OA) を用いた。

$$OA = \frac{TP + TN}{TP + TN + FP + FN}$$

ここでTPはtrue positive, TNはtrue negative, FPはfalse positive, FNはfalse negativeの物質数である。

### 3 結果と考察

89種の芳香族N-ニトロソ・ニトロソ・ニトロ化合物について、本研究の予測結果（モデルⅢ）を初報（Tanabe et al, 2010, Tanabe et al, 2011a）のモデルⅠと次報（Tanabe et al, 2011b; Suzuki et al, 2012）のモデルⅡと比較した（Table 3）。モデルⅠとⅡを比較すると、感度分析による有効記述子選定の効果により発ガン性予測の正解率は70.8%から77.5%に向上したが、モデルⅡに6種の電子記述子を追加したモデルⅢでは、正解率の上昇は全く認められなかった。

この電子記述子追加の有効性の理由について初報では、芳香族N-ニトロソ・ニトロソ・ニトロ化合物では置換基の位置や種類の違いにより電子状態が異なり、その違いによってこれまで予測できなかった化合物の発ガン性が予測可能にあることを予測した。しかし、Table 3のように、電子記述子の追加の効果がなかったことは、このような置換基による電子状態の違いが発ガン性には、ほとんど影響せず、モデルⅡの記述子セットでカバーされている分子の立体的な形状や大きさ、官能基の有無などの分子のバルクの特徴が支配的であることを示唆するものであると考えられる。事実、Table 1の電子記述子が発ガン性に寄与する重要度の順位はTable 4に示すようであり、いずれも寄与順位はきわめて低く、感度分析法で選定された有効記述子に含まれていない。

しかし、始めに記したように、電子記述子を用いた発ガン性予測に関するこれまでの研究では電子記述子の有効性が確認されている。それらの先行研究と、電子記述子の効果が見られなかった本稿との違いはどこにあるかを考察すると、次の結論にたどりつく。化学物質の発ガン性予測に関する研究は2種類に大別できる。第1は、上記の先行研究のように、発ガン性の有無ではなく、発ガン性の強度（半数致死量LD<sub>50</sub>、等）のデータを用いて、発ガン性強度を予測するモデルを検証する研究である。これらの発ガン性強度予測研究では、分子の大きさや形状がほとんど同じ一連の同族体の内で発ガン性データのある物質のみを用いて電子記述子と発ガン性強度との関係を解析する。この関係は、QSARの原理である置換基定数と化学反応速度との直線的自由エネルギー関係を示すHammett則に対応させると、生体内での化学物質と生体側のレセプターとの化学反応速度が大きいほど発ガン性の強度が高くなることになるので、発ガン性の強度と置換基定数の関係はHammett則で保証されていることになる。したがって、発ガン性の強度を予測するモデルにおいては電子記述子の有効性が高いこと、また、電子記述子を用いて発ガン性の強度を高い精度で予測するモデルの作成が比較的容易なことも理解できる。

一方、第2の研究は本稿のように、発ガン性の強度の大小に関係なく、発ガン性の有無に関するデータのある化学物質を用いて発ガン性の有無を予測するモデルを構築する研究である。既存化学物質の管理、あるいは新規化学物質の事前予測といった観点からは、発ガン性の強度でなく、この発ガン性の有無を予測するモデルの方が実用的である。しかし、発ガン性の有無と置換基効果との関係をQSARの原理である有機反応理論に対応させると、化学反応が起こるか否かを置換基定数で説明することになるが、化学反応の有無はHammett則では保証されていない。したがって、本研究のように発ガン性の有無を予測するモデル作成の困難さ、および、そのような予測モデルにおける電子記述子の有効性の低

さが理解できよう。したがって、本報で取り上げた芳香族N-ニトロソ・ニトロソ・ニトロ化合物の発ガン性予測性能の向上のためには、別のタイプの記述子を採用するなど新たな方策を講ずる必要がある。

## 4 結論

本研究では、芳香族N-ニトロソ・ニトロソ・ニトロ群の89分子について分子軌道法により電子状態を計算し、幾つかの電子状態を表す記述子を求めた。その記述子群を、これまでの我々の発ガン性予測モデルに追加して解析を行い、正解率が向上するか否かを検討した。その結果、この種の化合物群では、置換基による電子状態の違いによる発ガン性への影響は、ほとんど認められないことが明らかになった。

化学物質の発ガン性予測は既存化学物質の管理、あるいは新規化学物質の事前予測といった観点からは必要性がきわめて高い。しかし、著者らが以前に作成した発ガン性予測モデルの性能は先行研究を大幅に超える性能であるが、動物実験の代替としての発ガン性予測手法の正解率としては未だ満足できる数値ではない。したがって、今後、本手法による発ガン性予測手法の性能を向上させるためには、実験データの拡充、新規記述子の作成、新規解析手法の検討等、様々な手段を検討する必要がある。

## 謝 辞

本研究において協力いただいたたBono Lučić博士 (The Rudjer Bošković Institute, Croatia), Dragan Amić (The Josip Juraj Strossmayer University, Croatia) 博士に感謝します。

## 参考文献

- Benigni, R. (2005) Structure-Activity Relationship Studies of Chemical Mutagens and Carcinogens: Mechanistic Investigations and Prediction Approaches. *Chem. Rev.* **105**, 1767–1800.
- Bohari, M. H.; Srivastava, H. K. (2011) Analogue-based Approaches in Ant-cancer Compound Modelling : The Relevance of QSAR Models. *Org. Med. Chem. Lett.* **1**, 1–12.
- Contrera, J. F.; Matthews, E. J.; Benz, R. D. (2003) Predicting the Carcinogenic Potential of Pharmaceuticals in Rodents Using Molecular Structural Similarity and E-State Indices. *Regul. Toxicol. Pharmacol.* **38**, 243–259.
- Contrera, J. F.; Kruklak, N. L.; Matthews, E. J.; Benz, R. D. (2007) Comparison of MC4PC and MDL-QSAR Rodent Carcinogenicity Predictions and the Enhancement of Predictive Performance by Combining QSAR Models. *Regul.*

- Toxicol. Pharmacol.* **49**, 172–182.
- Fjodorova, N.; Novic, M. (2012) Integration of QSAR and SAR Methods for the Mechanistic Interpretation of Predictive Models for Carcinogenicity. *Res. Rev. Article*, **1**, e201207003, <http://dx.doi.org/10.5936/csbj.201207003>.
- Fioravanzo, E.; Bassan, A.; Pavan, M.; Mostrag-Szlichtyng, A.; Worth, A. P. (2012) Role of in silico Genotoxicity Tools in the Regulatory Assessment of Pharmaceutical Impurities. *SAR QSAR Environ. Res.* **23**, 257–277.
- Gini, G.; Lorenzini, M.; Benfenati, E.; Grasso, P.; Bruschi, M. (1999a) Predictive Carcinogenicity: A Model for Aromatic Compounds with Nitrogen-Containing Substituents Based on Molecular Descriptors Using an Artificial Neural Network. *J. Chem. Inf. Comput. Sci.* **39**, 1076–1080.
- Gini, G.; Lorenzini, M.; Vittore, A.; Benfenati, E.; Grasso, P. (1999b) Some Results for the Prediction of Carcinogenicity Using Hybrid Systems. In: *Predictive toxicology of chemicals: Experience and impact of AI tools*, AAI Press, Menlo Park, California, pp. 139–143.
- Greene, N. (2002) Computer Systems for the Prediction of Toxicity; An Update. *Adv. Drug Delivery Rev.* **54**, 417–431.
- Hemmateenejad, B.; Safarpour, M. A.; Miri, R.; Nesari, N. (2005) Toward Optimal Procedure for PC-ANN Model Building: Prediction of the Carcinogenic Activity of a Large Set of Drug. *J. Chem. Inf. Comput. Sci.* **45**, 190–199.
- Pullman, A. (1945) On an Electronic Theory of the Carcinogenic Action of Condensed Aromatic Hydrocarbons. *C. R. Séance Soc. Biol. Fil.* **139**, 1056–1058.
- Schmidt, O. (1939) Die Charakterisierung der einfachen und Krebs erzeugenden aromatischen Kohlenwasserstoffe durch die Dichteverteilung bestimmter Valenzelektronen. *Z. Physik. Chem.* **B42**, 83–110.
- Serafimova, R.; Gatnik, M. F.; Worth, A. (2010) Review of QSAR Models and Software Tools for Predicting Genotoxicity and Carcinogenicity. JRC Technical Report EUR 24427 EN, Publications Office of the European Union, Luxembourg.
- Suzuki, T.; Nishida, K.; Tanabe, K. (2012) QSAR Prediction of Carcinogenicity of Nitro and Related Compounds. *J. Toyo Univ. Natur. Sci.* **56**, 65–73.
- Tanabe, K.; Lučić, B.; Amić, D.; Kurita, T.; Kaihara, M.; Onodera, N.; Suzuki, T. (2010) Prediction of Carcinogenicity for Diverse Chemicals Based on Substructure Grouping and SVM Modeling. *Mol. Div.* **14**, 789–802.
- Tanabe, K.; Kurita, T.; Nishida, K.; Suzuki, T. (2011a) Construction of a Chemical Carcinogenicity Database and a Carcinogenicity Prediction System from Chemical Structures. *J. Toyo Univ. Natur. Sci.* **55**, 29–61.
- Tanabe, K.; Nishida, K.; Suzuki, T. (2011b) Prediction of Carcinogenicity of Chemical Substances with Support Vector Machine. *J. Comput. Chem. Jpn.* **10**, 115–121.
- Vendrame, R.; Braga, R. S.; Takahata, Y.; Galvao, D. S. (1999) Structure-

- Activity Relationship Studies of Carcinogenic Activity of Polycyclic Aromatic Hydrocarbons Using Calculated Molecular Descriptors with Principal Component Analysis and Neural Network Methods. *J. Chem. Inf. Comput. Sci.* **39**, 1094–1104.
- Vracko, M. (1997) A Study of Structure-Carcinogenicity Potency Relationship with Artificial Neural Networks. The Using of Descriptors Related to Geometrical and Electronic Structures. *J. Chem. Inf. Comput. Sci.* **37**, 1037–1043.