Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 916–934
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Comparative QSTR studies for predicting mutagenicity of nitro compounds

Pramod C. Nair, M. Elizabeth Sobhia *

Centre for Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, S.A.S Nagar, Punjab 160062, India

Received 5 February 2007; received in revised form 21 June 2007; accepted 25 June 2007
Available online 28 June 2007

Abstract

Mutagenicity and carcinogenicity are toxicological endpoints which pose a great concern being the major determinants of cancers and tumours. Nitroarenes possess genotoxic properties as they can form various electrophilic intermediates and adducts with biological systems. Different QSTR techniques were employed to develop models for the prediction of mutagenicity of nitroarenes using a diverse set of 197 nitro aromatic and hetero aromatic molecules. The 2D and 3D QSTR methods used for model development gave statistically significant results. The alignment for 3D methods was obtained by maximum common substructures (MCS) approach, by taking the most mutagenic molecule of the dataset as the template. All the QSTR models were developed with the same set of training and test set molecules. The 3D contours and 2D contribution maps along with molecular fingerprints provide useful information about the mutagenic potentials of the molecules. The GFA based model shows thermodynamic and topological descriptors play an important role in characterizing mutagenicity of nitroarenes. Atomic-level thermodynamic descriptor namely AlogP throws light on hydrophobic features and helps to understand the bilinear model. Topological aspects of these classes of compounds were depicted by the fragment fingerprints and Balaban indices obtained from HQSAR and GFA models, respectively. The predictive abilities of 2D and 3D QSTR models may be useful as a vibrant predictive tool to screen out mutagenic nitroarenes and design safer non-mutagenic nitro compounds.

Keywords: Mutagenicity; Nitroarenes; QSTR; CoMFA; CoMSIA; HQSAR; GFA

1. Introduction

Over the past two decades, there has been a vast increase in the development of SAR techniques. The linear free-energy approaches developed by Hansch and Free-Wilson have provided a fundamental scientific framework for the study of structure activity relationships. The QSAR techniques involve the use of 2D or 3D descriptors such as electron distribution, spatial disposition molecular volume, hydrophobicity, steric feature, solubility and ionization constants for developing robust models [1–3]. The well recognized 3D-QSAR methods, CoMFA and CoMSIA have received much attention in drug discovery scenario [4,5]. More recently, the QSAR techniques were extrapolated to predict the other end points like toxicity and pharmacokinetic properties based on the molecular structures and named as QSTR and QSPR analysis [6,7]. Developing QSTR models for toxic end points are encouraging in particular to overcome the time and cost associated with the animal studies for screening the new compounds. However, there are many difficulties one has to face while developing QSTR models, one being the diversity of the molecules and other being the varied mechanisms by which the molecules cause the toxicological effect. In addition, QSAR methods do not necessarily model interactions with receptor and do not highlight the exact mechanism involved in complex enzymatic steps observed in a toxic scenario. But they do give some information of the involved mechanism based on the molecular similarity and this was recently pointed out by Gieleciak and Polanski [8]. There are also other challenges exist which include predicting test and new molecules. The toxicological properties like hepatotoxicity, carcinogenicity and reproductive effects are mechanistically unclear leading to added complexity for prediction. Owing to these reasons there are not many successful QSTR models available in the literature. However,

* Corresponding author. Tel.: +91 172 2214683.
E-mail address: mesophia@niper.ac.in (M.E. Sobhia).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2007.06.006

there are more than a few QSTR models reported for predicting toxicity for individual chemical classes such as quinolines, nitroarenes, aromatic amines, triazenes, polycyclic aromatic hydrocarbons and lactones which undergo similar mechanisms to cause toxicity.

Mutagenicity and carcinogenicity are toxicological endpoints which pose a great concern being the major determinants of cancers and tumours [9]. Carcinogenic process involves one or more mutations showing a relationship with mutagenicity [10–12]. Zeiger performed an analysis on 224 chemicals and revealed that all nitro carcinogens were mutagenic [12] in behaviour. Several groups have used log revertants data from Ames test of mutagenicity for building QSTR models of mutagenic compounds [13–18]. The Ames test (*Salmonella typhimurium his* reversion assay) is a simple and less costly method for finding out mutagenic potential of the compounds. It is performed on different strains of *S. typhimurium* depending on the mechanism and metabolic activation of different classes of carcinogens [19–22]. Literature studies reveal that *S. typhimurium* TA98 undergoes mutations in presence of nitro compounds; therefore one can use the TA98 strain for the generation of QSTR models [2,23]. Chemical carcinogens are either genotoxic or non-genotoxic based on their mechanism of action [24]. Genotoxic carcinogens directly cause damage to DNA and many known mutagens belong to this category [25]. Non-genotoxic carcinogens do not bind covalently to DNA, so do not directly cause damage to the DNA.

We report here the results of the QSTR studies carried out on a dataset of 197 nitroarene compounds. Nitroarenes are environmental pollutants released from automobile exhausts and industrial areas, proved to be potent mutagens or carcinogens [26,27,30]. There are also examples in the literature to show drugs and drug-like molecules have nitro substitutions which may cause toxicological influence [28,29]. Nitroarenes possess genotoxic properties as they can form various electrophilic intermediates and adducts with DNA, tissue proteins, blood proteins albumin and haemoglobin [25]. The proposed mechanism for nitro compounds involves transformation of nitro into a hydroxylamine intermediate. This results in an electrophilic nitrogen species interacting with DNA which is mediated by cytosolic reductases. The objective of the present work is to develop comparative QSTR models using 2D and 3D QSAR methods to get better insight into the mutagenicity of nitro compounds. This study may help to understand the influence of fragments fingerprints and other physicochemical descriptors on the mutagenic potential of nitroarenes. This may also aid to analyse the effect of various fields of interactions of CoMFA and CoMSIA on mutagenicity of nitroarenes. Hansch et al. published a conventional QSAR study on series of nitroarenes; however, we have approached this problem with the more advanced 2D QSAR and 3D QSAR methods, viz. HQSAR, GFA, CoMFA and CoMSIA.

2. Dataset for analysis

Mutagenicity data on *S. typhimurium* TA98 strain in log revertants/nmol was used for the study. The average log revertants/nmol values for 197 nitroarenes from published data by Goldring et al., Rosenkranz et al., Zielinska et al. [31–33] and Hansch et al. where taken for developing QSAR model (Table 1) [34]. Reasons for considering average values were to reduce the experimental errors which could be because of

Table 1
Molecules in the dataset with their log revertant values used for QSTR analysis

S.no.	Molecules	Molecule name	log revertants (nmol)	Reference
1	**1**	2-Nitrophenanthrene	2.11	[32]
2	**2**	8-Nitroquinoline	−1.24	[34]
3	**3**	5-Nitroquinoline	−0.96	[34]
4	**4**	1-Nitrofluoranthene	2.74	[32]
5	**5**	2,5-Difluoronitrobenzene	−0.79	[34]
6	**6**	1,3,6,8-Tetranitropyrene	4.99	[32]
7	**7**	3-Methyl-4-nitrobiphenyl	−0.10	[34]
8	**8**	2-Acetoxy-7-nitrofluorene	1.86	[34]
9	**9**	5-Nitrobenzimidazole	−1.83	[34]
10	**10**	2-Methyl-7-nitrofluorene	2.36	[34]
11	**11**	9-Nitrophenanthrene	2.25	[34]
12	**12**	1-Amino-8-nitropyrene	2.43	[34]
13	**13**	2,4-Dinitrofluoranthene	3.78	[33]
14	**14**	1-Methyl-6-nitroindazole	−1.10	[34]
15	**15**	1-Nitropyrene	2.78	[32]
16	**16**	3,4-Dinitrofluoranthene	3.62	[34]
17	**17**	2-Nitro-*m*-phenylenediamine	−3.00	[34]
18	**18**	3-Nitro-9-fluorenone	2.13	[32]
19	**19**	6-Nitroquinoline	−1.08	[34]
20	**20**	2-Cyano-7-nitrofluorene	2.51	[34]
21	**21**	4,3′-Dinitrobiphenyl	0.23	[34]
22	**22**	2-Iodo-7-nitrofluorene	2.97	[34]
23	**23**	2-Nitrobenz(*i*)acenthrylene	0.86	[31]
24	**24**	1-Nitrocoronene	0.45	[34]
25	**25**	2-Hydroxy-7-nitrofluorene	1.68	[34]

Table 1 (*Continued*)

S.no.	Molecules	Molecule name	log revertants (nmol)	Reference
26	**26**	1-Chloro-2,4-dinitrobenzene	0.30	[34]
27	**27**	1-Nitrobenzo[*a*]pyrene	2.63	[34]
28	**28**	2-Fluoro-7-nitrofluorene	2.68	[34]
29	**29**	2,4,7-Trinitro-9-fluorenone	3.41	[32]
30	**30**	4-Nitrobenz[*k*]acenthrylene	0.67	[31]
31	**31**	1,2-Dinitrofluoranthene	3.11	[33]
32	**32**	2,4,3′,4′-Tetranitrobiphenyl	1.54	[34]
33	**33**	2-[(Trifluoroacetyl)amino]-7-nitrofluorene	2.81	[34]
34	**34**	2-Nitronapthalene	−0.30	[34]
35	**35**	2-Amino-7-nitrofluorene	1.56	[32]
36	**36**	3-Methyl-2-nitonaphthalene	0.00	[34]
37	**37**	4-Nitrostilbene	0.69	[34]
38	**38**	4-Nitrobiphenyl	−0.30	[34]
39	**39**	6-Mithoxy-8-nitroquinoline	−1.21	[34]
40	**40**	1-Nitronaphthalene	−0.61	[32]
41	**41**	2-Bromo-7-nitrofluorene	3.06	[34]
42	**42**	2-Methyl-5-nitroindazole	−1.10	[34]
43	**43**	2-Nitrofluoranthene	3.01	[33]
44	**44**	2-Nitro-9,10-dihydrophenanthrene	1.99	[34]
45	**45**	1,4-Dinitrobenzene	0.15	[34]
46	**46**	1-Nitro-7,8,9,10-tetrahydrobenzo[*a*]pyrene	0.90	[34]
47	**47**	2,7-Dinitro-4,5,9,10-tetrahydropyrene	3.50	[34]
48	**48**	2-Nitroanthracene	2.95	[32]
49	**49**	2,8-Dinirophenazine	2.75	[34]
50	**50**	1,5-Dinitronaphthalene	0.52	[34]
51	**51**	1,6-Dinitro-9,10,11,12-tetrahydrobenzo[*e*]pyrene	2.41	[34]
52	**52**	1,3-Dinitro-9,10,11,12-tetrahydrobenzo[*e*]pyrene	2.41	[34]
53	**53**	2,7-Dinitrofluorene	3.22	[32]
54	**54**	7-Nitro-2,3-dichlorodibenzo-1,4-dioxin	1.73	[34]
55	**55**	1-Methyl-2-nitronaphthalene	−0.70	[34]
56	**56**	2-Nitrophenazene	2.06	[34]
57	**57**	2,4,4′-Trinitrobiphenyl	0.66	[34]
58	**58**	*trans*-9,10-Dihydro-3-nitrobenzo[*a*]pyrene-9,10-diol	2.80	[34]
59	**59**	2-Nitrobenz[l]acenthrylene	0.26	[31]
60	**60**	1,3,8-Trinitronaphthelene	0.35	[34]
61	**61**	2-Chloro-7-nitrofluorene	3.11	[34]
62	**62**	2-Amino-5-nitrophenol	−2.40	[34]
63	**63**	*N*′-(5-nitro-2-furfuryliden)-5-nitro-2 furanacrylohydrazide	2.45	[34]
64	**64**	2,7-Dinitro-9-fluorenone	3.19	[32]
65	**65**	1,3-Dinitronaphthalene	−0.05	[34]
66	**66**	4-Amino-3′-nitrobiphenyl	−1.52	[34]
67	**67**	2-Hydroxy-1-nitrofluoranthene	2.26	[33]
68	**68**	3-Nitrobenzo[*a*]pyrene	2.82	[34]
69	**69**	2,5-Dinitrofluorene	3.20	[34]
70	**70**	5-Nitroisatin	−0.94	[34]
71	**71**	1,8-Dinitro-9,10,11,12-tetrahydrobenzo[*e*]pyrene	2.19	[32]
72	**72**	2-Nitrofluorene	1.43	[34]
73	**73**	3-Amino-2′-nitrobiphenyl	−2.00	[34]
74	**74**	3-Nitrobenzo[*k*]fluoranthene	2.76	[34]
75	**75**	4,4′-Dinitrobiphenyl	1.17	[34]
76	**76**	2-Chloronitrobenzene	−1.72	[34]
77	**77**	2-Amino-3-nitrobiphenyl	−1.52	[34]
78	**78**	2′-Methyl-4-nitrobiphenyl	−0.23	[34]
79	**79**	2-Nitrodibenzo-1,4-dioxin	1.79	[34]
80	**80**	7-Nitrofluoranthene	1.87	[32]
81	**81**	1,3-Dinitrobenzene	0.03	[34]
82	**82**	1,2,4-Trinitrofluoranthene	3.56	[33]
83	**83**	2-Nitrocarbazole	1.01	[34]
84	**84**	4-Nitrostyrene	−1.30	[34]
85	**85**	6-Nitrobenz[*e*]acenthrylene	0.04	[31]
86	**86**	5-Nitrobenz[*k*]acenthrylene	0.92	[31]
87	**87**	2-Nitrodibenzofuran	1.64	[34]
88	**88**	4-Nitroacetophenone	−1.54	[34]
89	**89**	1-Methyl-7-nitroindazole	−1.00	[34]
90	**90**	1-Nitrobenzo[*e*]pyrene	1.59	[34]

Table 1 (*Continued*)

S.no.	Molecules	Molecule name	log revertants (nmol)	Reference
91	**91**	*trans*-7,8-Dihydro-3-nitrobenzo[*a*]pyrene-7,8-diol	3.08	[34]
92	**92**	2-Nitro-4,5,9,10-tetrahydropyrene	1.58	[34]
93	**93**	3-Nitrobenzo[e]pyrene	2.95	[34]
94	**94**	4-Nitropyrene	3.39	[34]
95	**95**	2,3,5-Trinitronaphthalene	1.51	[34]
96	**96**	1,3,6-Trinitropyrene	4.99	[32]
97	**97**	2-Methoxy-7-nitrofluorene	2.79	[34]
98	**98**	5-Nitroisoquilonene	−1.55	[34]
99	**99**	1-Nitroacenaphthylene	1.77	[34]
100	**100**	3-Amino-3′-nitrobiphenyl	−1.70	[34]
101	**101**	3-Nitro-9,10,11,12-tetrahydrobenzo[e]pyrene	0.78	[34]
102	**102**	2,4,5,7-Tetranitro-9-fluorenone	2.93	[34]
103	**103**	2-Nitropyrene	3.35	[34]
104	**104**	1-Nitro-9,10,11,12-tetrahydrobenzo[*e*]pyrene	0.70	[34]
105	**105**	5-Nitro-1,10-phenanthroline	0.59	[34]
106	**106**	10-Hydroxy-1-nitropyrene	1.89	[34]
107	**107**	2,3,5-Trinitrofluoranthene	3.44	[33]
108	**108**	3-Nitrofluoranthene	3.67	[32]
109	**109**	5,10-Dinitrobenzo[*ghi*]perylene	3.60	[34]
110	**110**	2,4,2′,6′-Tetranitrobiphenyl	−0.07	[34]
111	**111**	4-Nitrobenzaldehyde	−1.64	[34]
112	**112**	3-Nitroacenaphthalene	1.00	[34]
113	**113**	2,4-Dinitroaniline	−1.74	[34]
114	**114**	2-Amino-4′-nitrobiphenyl	−1.70	[34]
115	**115**	5-Nitroindene	0.08	[34]
116	**116**	4-Fluoronitrobenzene	−0.23	[34]
117	**117**	2,4,3′-Trinitrobiphenyl	0.03	[34]
118	**118**	3-Nitro-7,8,9,10-tetrahydrobenzo[*a*]pyrene	0.30	[34]
119	**119**	2,4,2′-Trinitrobiphenyl	−0.19	[34]
120	**120**	1-Methyl-5-nitroindazole	−0.82	[34]
121	**121**	1,2,5-Trinitrofluoranthene	3.16	[33]
122	**122**	6-Nitrochrysene	1.75	[34]
123	**123**	3,2′-Dimethyl-4-nitrobiphenyl	−0.84	[34]
124	**124**	1-Nitro-2,4-difluorobenzene	−1.66	[34]
125	**125**	2-Nitro-4,5-dihydropyrene	3.27	[34]
126	**126**	1,7-Sinitrophenazine	2.02	[34]
127	**127**	3,4,3′-Trinitrobiphenyl	1.92	[34]
128	**128**	5-Nitroacenaphthylene	1.91	[34]
129	**129**	2,4-Dinitrophenylhydrazine	−0.07	[34]
130	**130**	2-Nitroanisole	−2.70	[34]
131	**131**	4-Nitrochalcone	−1.15	[34]
132	**132**	2-Nitrobenzaldehyde	−1.92	[34]
133	**133**	2-Nitro-1,3,7,8-tetrachlorodibenzo-1,4-dioxin	−1.40	[34]
134	**134**	2,7-Dinitro-4,5-dihydropyrene	4.25	[34]
135	**135**	3-Chloro-4-fluronitrobenzene	−1.21	[34]
136	**136**	1-Hydroxy-3-nitropyrene	3.87	[32]
137	**137**	2,3-Dinitrofluoranthene	2.62	[33]
138	**138**	2-Methyl-6-nitroindazole	−0.41	[34]
139	**139**	4-Chloro-2-nitroaniline	−2.00	[34]
140	**140**	1,8-Dinitronaphthalene	0.90	[34]
141	**141**	2,3-Dichloronitrobenzene	−1.51	[34]
142	**142**	8-Nitro-2,3,7-trichlorodibenzo-1,4-dioxin	−0.53	[34]
143	**143**	2-Nitrophenetole	−2.22	[34]
144	**144**	2,5-Dichloronitrobenzene	−1.54	[34]
145	**145**	3-Amino-4′-nitrobiphenyl	0.25	[34]
146	**146**	4-Amino-4′-nitrobiphenyl	0.19	[34]
147	**147**	2-Methyl-5-nitrobenzimidazole	−0.51	[34]
148	**148**	3-Nitroacenaphthylene	1.77	[34]
149	**149**	2,7-Dinitrophenazine	4.34	[34]
150	**150**	Metronidazole	−1.61	[34]
151	**151**	7-Nitroindazole	0.11	[34]
152	**152**	2,5-Dinitrofluoranthene	2.32	[33]
153	**153**	5-Nitroacenaphthene	0.58	[32]
154	**154**	2-Nitrobenzimidazole	0.00	[34]
155	**155**	2,4-Dinitroanisole	−1.89	[34]

Table 1 (*Continued*)

S.no.	Molecules	Molecule name	log revertants (nmol)	Reference
156	**156**	6-Nitroindoline	−0.48	[34]
157	**157**	6-Hydroxy-1-nitropyrene	1.34	[34]
158	**158**	2-(Acetylamino)-7-nitrofluorene	2.85	[34]
159	**159**	8-Hydroxy-1-nitropyrene	1.49	[34]
160	**160**	2-Bromo-4,6-dintroaniline	−1.32	[34]
161	**161**	1,6-Dinitrobenzo[*e*]pyrene	1.99	[34]
162	**162**	3,4,3′,4′-Tetranitrobiphenyl	2.85	[34]
163	**163**	2,7-Dinitropyrene	4.58	[34]
164	**164**	1,3-Dinitropyrene	5.04	[32]
165	**165**	8-Nitrofluoranthene	4.05	[32]
166	**166**	1-Nitro-3-acetoxypyrene	4.22	[34]
167	**167**	1,6-Dinitropyrene	5.06	[32]
168	**168**	4-Nitroanisole	−2.70	[34]
169	**169**	2,4-Dinitro-1-fluorobenzene	1.20	[34]
170	**170**	1-Nitrophenazine	0.87	[34]
171	**171**	2,4,2′,4′-Tetranitrobiphenyl	4.58	[34]
172	**172**	6-Nitroindazole	0.66	[34]
173	**173**	3,4,4′-Trinitrobiphenyl	2.60	[34]
174	**174**	3,9-Dinitrofluoranthene	5.02	[34]
175	**175**	3-nitro-1,2,4,7,8-pentachlorodibenzo-1,4-dioxin	−0.33	[34]
176	**176**	6-Nitroindene	0.96	[34]
177	**177**	1,9-Dinitrophenazine	1.26	[34]
178	**178**	2-Methyl-7-nitroindazole	0.23	[34]
179	**179**	4,4′-Dinitrochalcone	−1.42	[34]
180	**180**	5,8-Dinitrobenzo[ghi]perylene	4.33	[34]
181	**181**	2-Nitro-*p*-phenylenediamine	−1.11	[34]
182	**182**	8-Nitroquinaldine	−2.70	[34]
183	**183**	1,8-Dinitropyrene	5.39	[32]
184	**184**	3,7-Dinitrofluoranthene	5.09	[34]
185	**185**	3,4-Dintro-1-fluorobenzene	−1.84	[34]
186	**186**	5-Nitroindoline	−0.17	[34]
187	**187**	2-Nitrotriphenylene	4.09	[34]
188	**188**	3-Nitrocarbazole	−0.70	[34]
189	**189**	1,3,6,8-Tetranitronaphthalene	−0.70	[34]
190	**190**	5-Nitroindole	0.57	[34]
191	**191**	4-Nitropenta[*cd*]pyrene	0.77	[31]
192	**192**	5-Nitrochrysene	-0.22	[34]
193	**193**	2-Nitrochrysene	−0.22	[34]
194	**194**	4-Nitro-*o*-phenylenediamine	0.63	[34]
195	**195**	1,2,3-Trichloro-4-nitrobenzene	−2.94	[34]
196	**196**	4-Nitrocarbazole	−2.00	[34]
197	**197**	1-Methyl-2-nitrobenzimidazole	2.54	[34]

variation in assay techniques or laboratory conditions. Dataset of molecules used for this study possess diverse structural variation ranging from a single benzene ring to coronene which consists of six fused benzene rings. The dataset includes around 34 nitropyrenes derivatives. The training and the test set were scaled randomly with **142** and **55** molecules in training and test set, respectively. Fig. 1 shows templates representing different classes of the molecules used in this study.

3. Computational details

Molecular modeling studies were performed using the molecular modeling package SYBYL7.1 [35] and Cerius^2 4.9 [36] installed on a Silicon Graphics Fuel Workstation. The structures were minimized using Powell's conjugate gradient method with Tripos force field and Gasteiger-Huckel partial atomic charges [37,38]. The minimum energy difference of 0.05 kcal/mol was set as a convergence criterion.

3.1. Alignment rule

The molecules in the dataset were aligned using nitro substituted aromatic ring systems. Rigid alignment was achieved due to presence of rigid aromatic rings present in the database using Distill module of Sybyl [35] The molecule **183** was taken as the template. The rigid alignment attempts to align molecules in a database to a template molecule on a common backbone or core. The minimum atom count in maximum common substructures (MCS) fragments can be as small as three. Mapping of core atoms is achieved by examining the RMS deviation between template and the molecule to be aligned and the lowest RMS deviation is chosen. For more than one mapping with the same RMS deviation the mapping that produces the lowest differential volume between the molecule and the template is chosen. In the present study, a single mapping was adopted where the molecules were fitted to the template using the best mapping of the core to the molecule.

The database was updated with the molecule's new orientation. Makhija et al. and Raichurkar et al. adopted maximum common substructure (MCS) alignment for CoMFA and CoMSIA analysis [39,40]. The aligned molecules of training set are shown in Fig. 2.

3.2. CoMFA and CoMSIA

CoMFA samples the steric and electrostatic fields surrounding a set of ligands and constructs a 3D-QSAR model by correlating these 3D fields with the corresponding biological/experimental activities [41]. The steric and electrostatic CoMFA potential fields were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å. The grid box dimensions were determined automatically in such a way that the region boundaries were extended beyond 4 Å in each direction from the co-ordinate of each molecule. The van der Waals potential and Coulombic terms, which represent steric and electrostatic fields, respectively, were calculated using the standard Tripos force fields. A distance dependent dielectric constant of 1.00 was used. An sp^3 hybridised carbon atom with +1 charge served as probe atom to calculate steric and electrostatic fields. The steric and electrostatic contributions were truncated to +30.0 kcal/mol and electrostatic contributions were ignored at the lattice intersections with maximal steric interactions.

The reported CoMSIA method is based on molecular similarity indices [42]; this method overcomes some of the drawbacks arising from the functional form of the Lennard–Jones and Coulomb potentials used in CoMFA. Molecular similarity is expressed in terms of five different properties, viz. steric, electrostatic, hydrophobic, H-bond donors and acceptors which were calculated using a C^+ probe atom with a radius of 1 Å placed at a regular grid spacing of 2 Å. CoMSIA similarity indices (AF) for molecule j with atoms i at a grid point q are calculated by the equation

$$A^q_{F,k}(j) = -\Sigma w_{\text{probe},k} w_{ik} \mathrm{e}^{-^{\text{ar}} iq^2}$$

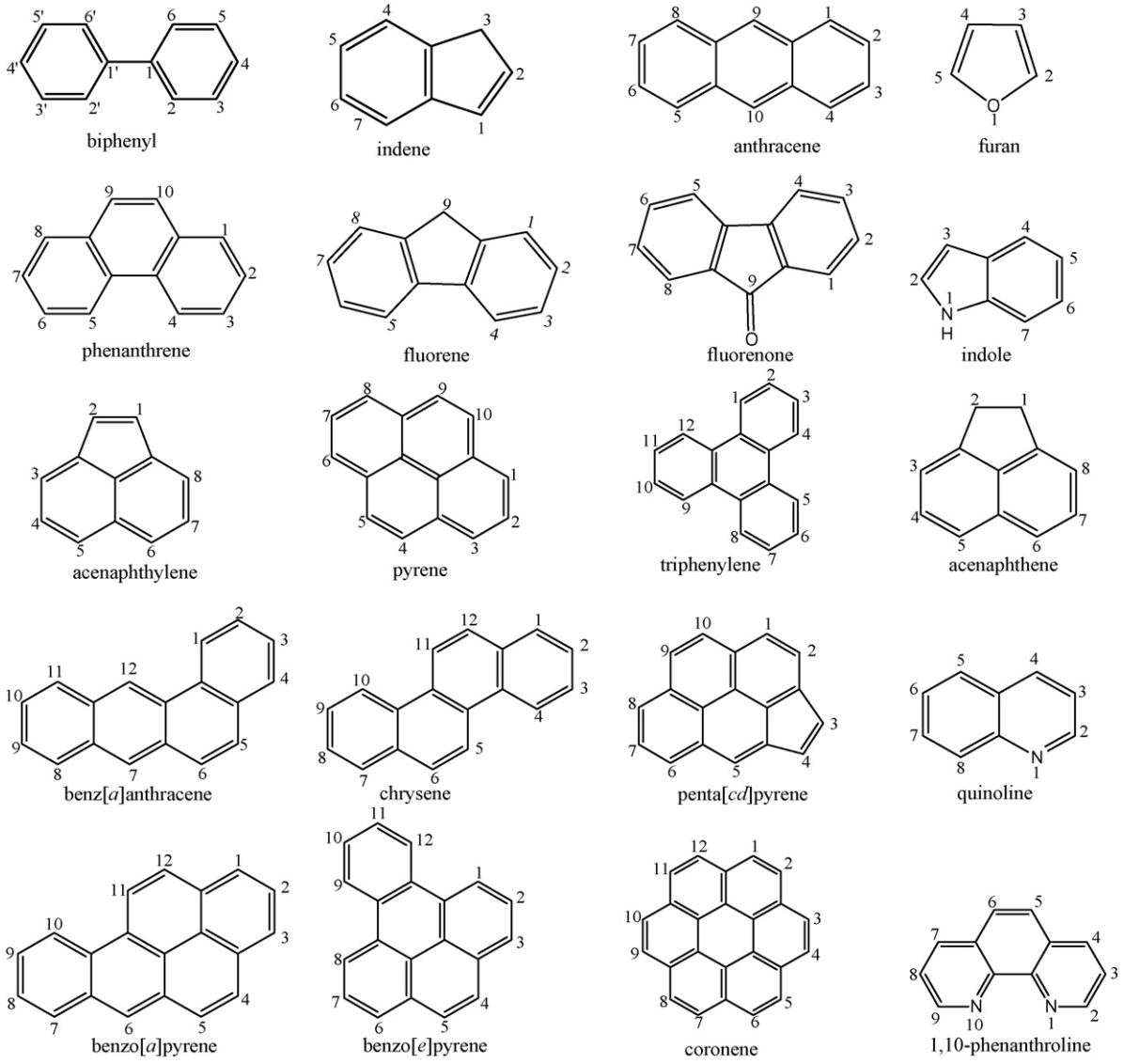

Fig. 1. Templates representing different classes of the molecules in the dataset.

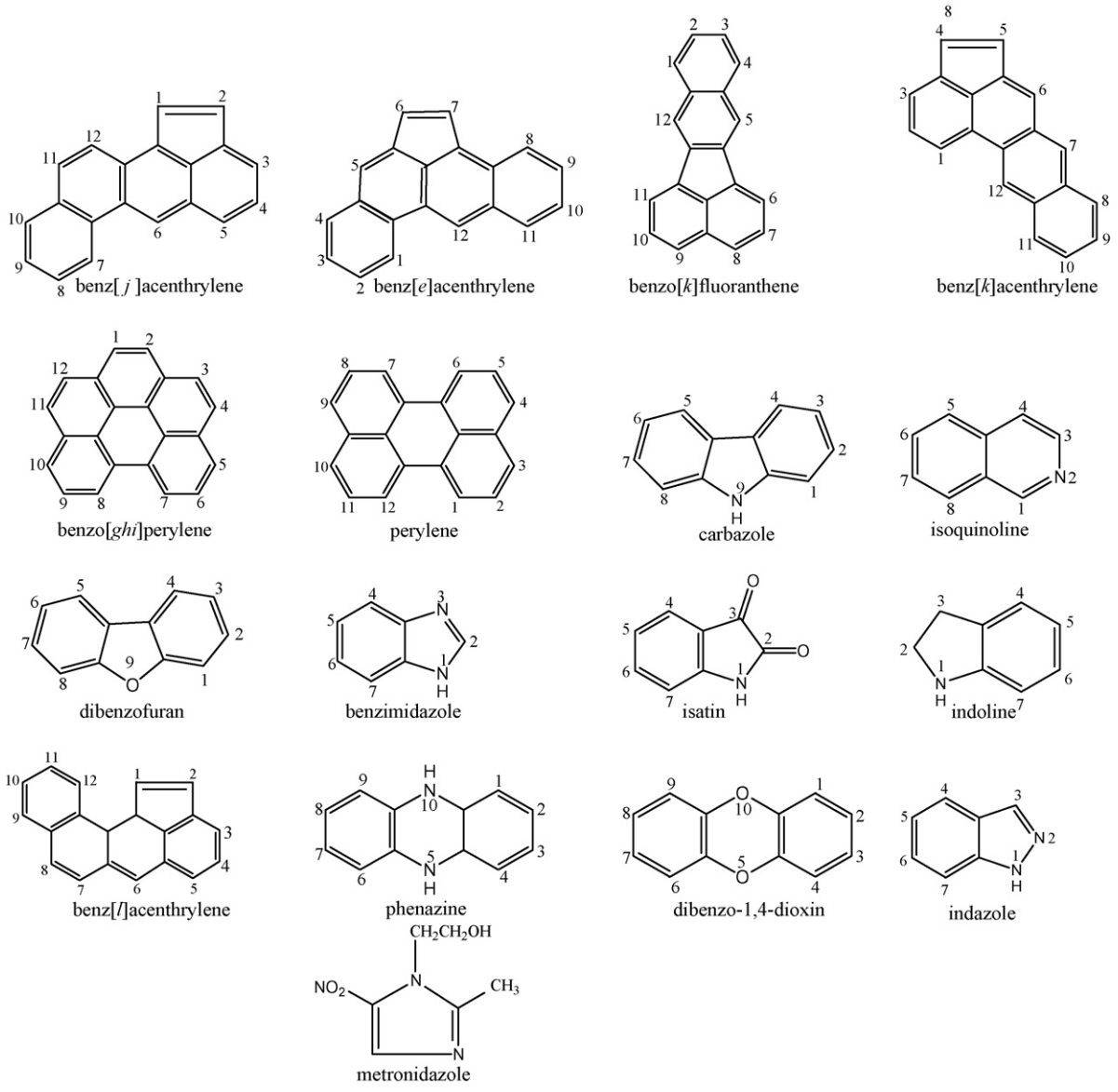

Fig. 1. (*Continued*).

where k represents the following physicochemical properties: steric electrostatic, hydrophobic, H-bond donor and H-bond acceptor. A Gaussian type distance dependence was used between grid point q and each atom i of the molecule. The default value of 0.3 was used as the attenuation factor (a). The steric indices are related to the third power of the atomic radii, electrostatic descriptors are derived from atomic partial charges, hydrophobic fields are derived from atom-based parameters [43] and H-bond donor and acceptor indices are obtained by a rule based method based on experimental results [44].

3.3. HQSAR

Hologram QSAR (HQSAR) is a technique which employs fragment fingerprints or molecular holograms as predictive variables of biological activity or other structural related data [35]. Ignoring hydrogen atoms fragments were generated of different sizes and hashed into bins 1–85 of the fingerprint. The fragment size 2–5 atoms were considered best in the present study. This molecular fingerprint was broken into strings at fixed interval as specified by a hologram length (HL) parameter. The hologram length determines the number of bins in the hologram into which the fragments are hashed. Each corresponding fragment SLN is then mapped to a pseudo-random integer in the range 0–2 [31] using the cyclic redundancy check (CRC) algorithm. The integer generated by the CRC algorithm is unique and reproducible for each and every unique SLN string. The hashing then occurs by folding the pseudorandom integer for a particular SLN string into the bin range defined. In Hologram QSAR, bins contain information about the number of fragments hashed into each bin. The optimal HQSAR model was derived from screening through the 12 default HL values, which were a set of 12 prime numbers ranging from 53 to 401.

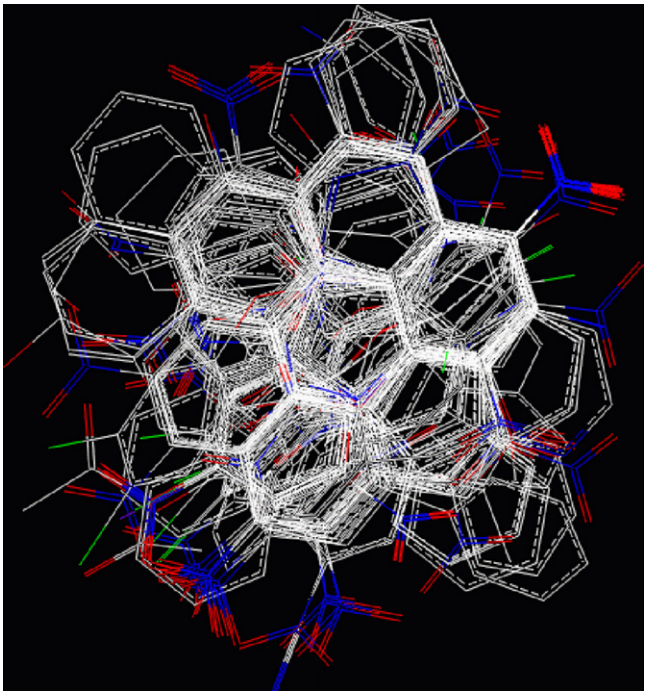

Fig. 2. Aligned molecules of the training set.

3.4. GFA based QSAR

Different descriptor classes such as E-state indices, electronic, spatial, structural, thermodynamic, information and topological descriptors were calculated with Cerius^2 4.9 package for the molecules in the dataset [45–47]. One hundred and forty descriptors were calculated for all the molecules in the dataset. Out of 140 descriptors 111 descriptors were used for model development. The remaining descriptors which did not show any contribution were rejected form the analysis. The inter correlation of descriptors was taken in to account and correlated descriptors were grouped together. Descriptors with highest correlation with log revertants were taken from the group (Supplementary data Table 1S).

The preferred descriptors were subjected to genetic function approximation (GFA) for selection of variables to obtain the QSAR models. GFA is genetics based method of variable selection, which combines Holland's genetic algorithm (GA) with Friedman's multivariate adaptive regression splines (MARS) [48,49]. The GFA method works by generating equations randomly. A default of 100 equations was used in the study. Then pairs of "parent" equations are chosen for "crossover" operations from this set of 100 equations randomly. The number of crossing over was set by default at 5000. The goodness of each progeny equation is assessed by Friedman's lack of fit (LOF) score, which is described by following formula

$$\mathrm{LOF} = \frac{\mathrm{LSE}}{\{1 - (c + dp)/m\}^2}$$

where LSE is the least-squares error, c the number of basis functions in the model, d the smoothing parameter, p the number of descriptors and m is the number of observations in the training set [48]. The smoothing parameter that controls the scoring bias between equations of different sizes was set at default value of 1.0 and the new term was added with a probability of 50%. Only the linear equation terms were used for model building. One hundred QSAR equations were generated using MARS and the best equation was selected on the statistical parameters such as regression coefficient, adjusted regression coefficient, regression coefficient cross validation and F-test values.

4. Results and discussion

4.1. CoMFA and CoMSIA

CoMFA and CoMSIA analysis were performed for a dataset comprising 197 nitroarenes. The dataset contains two to six fused member ring systems making manual alignment difficult and selection of template structure more critical. So the molecules were aligned using maximum common substructure (MCS), based on the template molecule **183**. A statistically significant model was obtained for the dataset containing **142** and **55** molecules in the training and test set, respectively. Table 2 shows the statistical parameters obtained from CoMFA and CoMSIA analysis. Tables 6 and 7 show the actual and predicted values from CoMFA study for the training and test set, respectively. Various CoMSIA models were tried considering steric, electrostatic, hydrogen donor, hydrogen acceptor and hydrophobic fields separately and in different combinations. CoMFA and CoMSIA 3D coefficient contour maps are shown in Figs. 3a and b and 4a–e, respectively, with the most mutagenic molecule **149** displayed in the background. The beneficial steric interaction regions are displayed by green contours while the detrimental ones are by yellow (Figs. 3 and 4a). The desirable positive and negative electrostatic interactions are illustrated by blue and red contours, respectively (Figs. 3 and 4b). Exact pinpointing of the various groups becomes difficult due to the diversity of the set and contour interpretation becomes futile when one thinks of more specific substitution. So, a general interpretation was drawn from the CoMFA and

Table 2
Statistical parameters obtained from CoMFA and CoMSIA analysis

	Statistical parameters	
	CoMFA	CoMSIA
r^2_{cv} (LOO)	0.760	0.710
NOC	15	15
r^2_{ncv}	0.986	0.968
F values	607.920	255.190
S.E.E.	0.235	0.359
Relative contributions		
Steric	0.429	0.240
Electrostatic	0.571	0.313
Hydrophobic	–	0.147
Donor	–	0.144
Acceptor	–	0.156

r^2_{cv} (LOO), Cross-validated correlation coefficient by leave one out; NOC, number of components; r^2_{ncv}, non-cross-validated correlation coefficient; S.E.E., standard error of estimate.

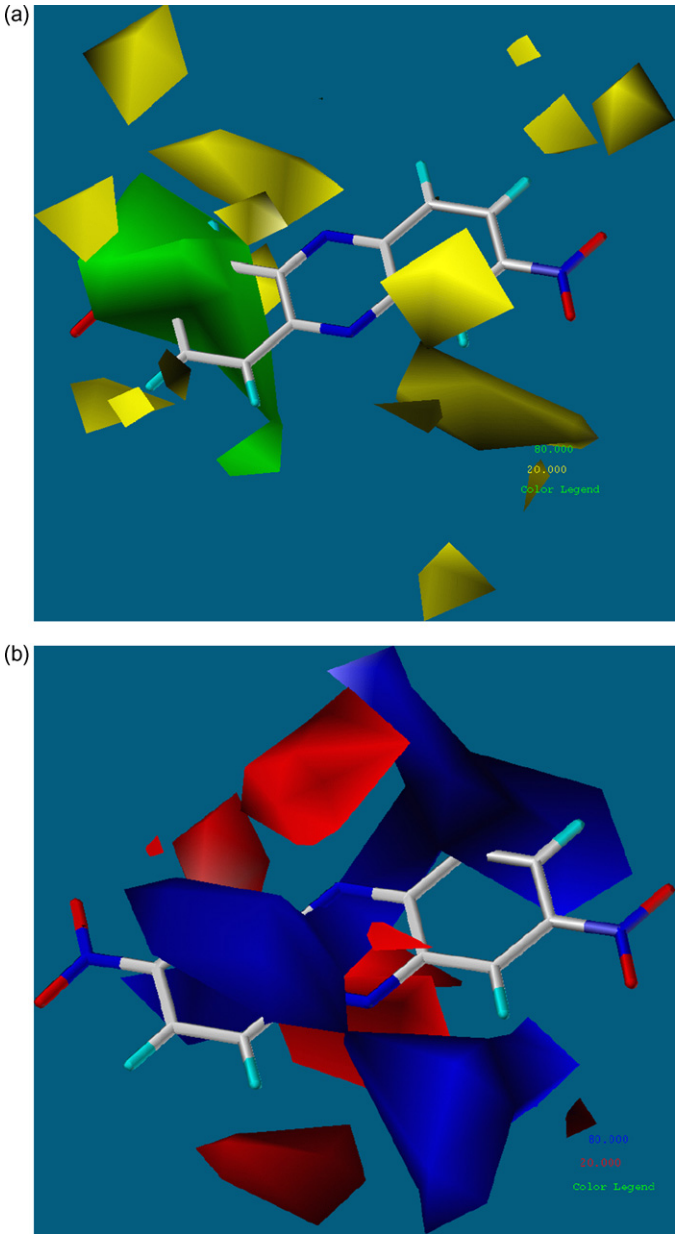

Fig. 3. (a) CoMFA STDEVXCOEFF steric contour maps. The most mutagenic molecule **149** of the training set is displayed in the background. (b) CoMFA STDEVXCOEFF electrostatic contour maps. The most mutagenic molecule **149** of the training set is displayed in the background.

CoMSIA study. Many of the potent mutagens show one of the rings penetrating green contour suggesting, the bulky substitutions may raise the mutagenicity of the molecules. This implies the presence of bulky substitutions enable the molecules to form stable complexes with DNA making them potential mutagens. Literature studies also showed the planarity of the molecules plays a key role in forming stable mutagen-DNA complex. And in particular the substitutions of bulky groups like phenyl may alter the planarity of the molecule, making the intercalation with DNA difficult. In such cases, the molecules may turn out to be less mutagenic. This phenomenon was observed in most of the biphenyls and allied systems [50–52]. On the other hand as noticed in nitropyrene derivatives, increase in size of fused rings in green contours without altering the planarity may improve mutagenicity (Fig. 2). In CoMFA bits of yellow contours are scattered around the molecule (Fig. 3a), while in CoMSIA they are found to be continuous (Fig. 4a). These contours show the lighter groups are favored at these regions for increasing mutagenicity. Many of the less mutagenic molecules possess smaller ring systems and are far away from the green contours. Also, the steric bulk in these mutagens is less in comparison to the potent mutagens. The electrostatic contour maps are shown in Figs. 3 and 4b. Most of the molecules partially embed their rings in red and blue regions. Dispersion of red contours around the molecule **149** may be due to the presence of mono, di, tri and tetra substituted nitro compounds in the dataset (Figs. 3 and 4b).

Studies reveal nitro group needs prior activation by nitro-reductases subsequently undergoes reduction via electron transfer process to generate a hydroxylamine intermediate. This reductive process involves gain of electron which depends on electronic charge of the molecule. The nitro group remains electronically rich when bonded to aromatic ring but in the presence of ortho substituted electron withdrawing or donating groups; it becomes electron deficient or rich. In other words, the electron withdrawing substituent in the vicinity of nitroarene pulls the electrons towards itself making nitro group electron deficient while electron donating substituent makes it electron rich. The advantage of CoMSIA analysis over CoMFA is that one can consider hydrophobic, hydrogen bond donor and acceptor parameters along with the steric and electrostatic ones.

Considering the hydrophobic contours many potent mutagenic molecules insert part of their rings in the hydrophobic favored yellow regions (Fig. 4c). The reverse is observed in the non-mutagenic and less mutagenic ones with rings positioned in the disfavored white region. Hansch et al. gave a major emphasis on hydrophobic ($\log P$) and electronic factors (ε_{LUMO}) in one of the QSAR studies and suggested that steric factors govern enzymatic activation for the compounds. However, their QSAR equation failed to highlight this point [34,53,54]. They also suggested a bilinear model where the hydrophobicity ($\log P$) increases with mutagenic potential but after certain point it shows decline suggesting requirement of optimum hydrophobicity. Mechanisms of mutagenesis by aromatic nitro compounds are complex and involve many steps of enzymatic activation to convert them into esters and/or nitrenium ion in their final step which may attack at several sites on DNA. The hydrogen bond donor and acceptor contour maps are shown in Fig. 4d and e, respectively. The high mutagenic molecules show the NO_2 groups seen in the favored magenta contour while low mutagenic ones in the disfavored red contour. As pointed out by Hansch et al. hydrophobicity is one of the important factors for a mutagen for its penetration into cell and the random walk through it. During the individual model development stage, it was seen that acceptor field depicts a high correlation with the mutagenicity of nitroarenes. Overall, one can say the hydrophobicity, steric, electronic and hydrogen acceptor parameters are important when a molecule has to tackle enzymes in every enzymatic step. However, while interacting with DNA, hydrophobicity plays a lesser role than

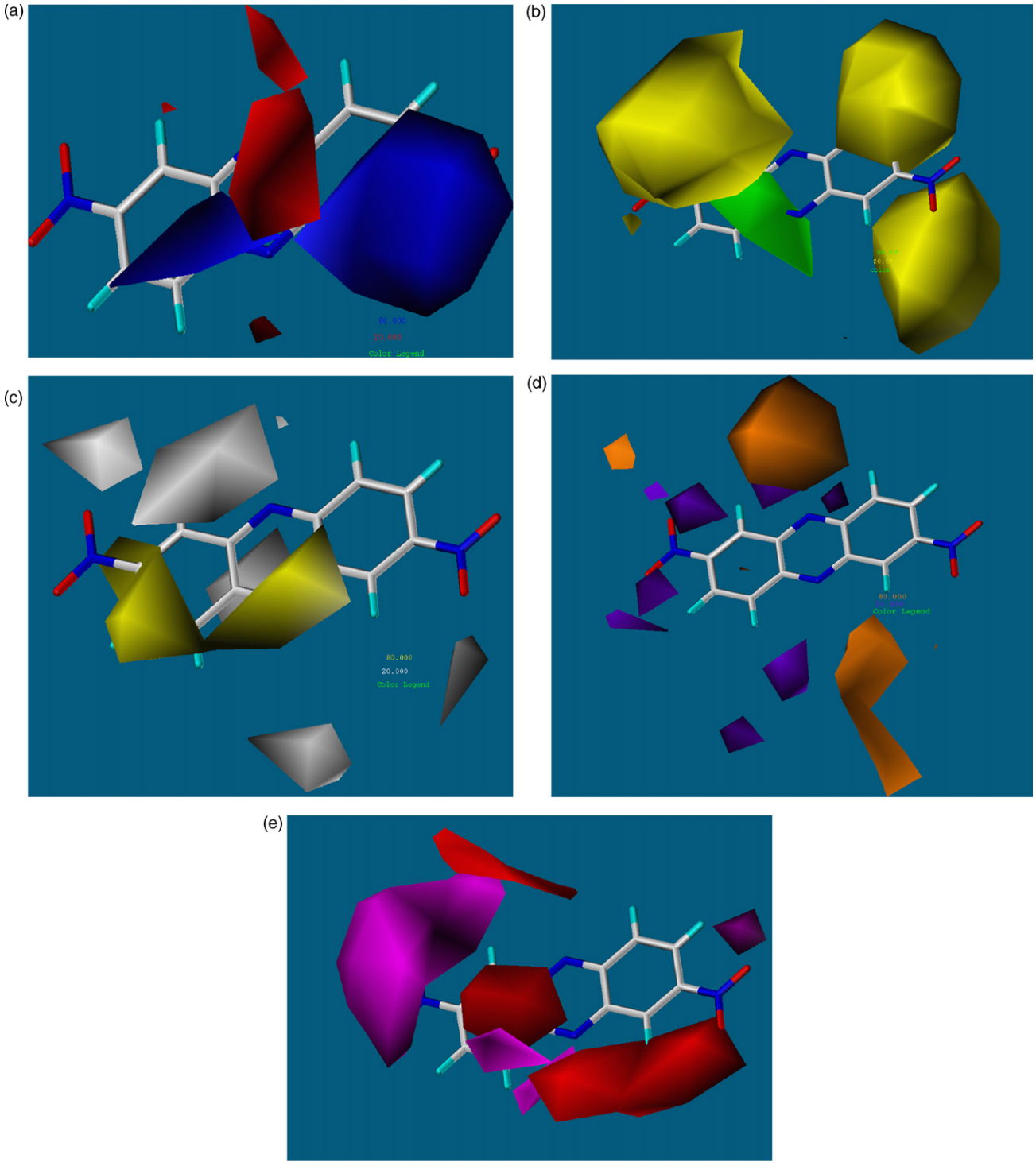

Fig. 4. (a) CoMSIA STDEVXCOEFF steric contour maps. The most mutagenic molecule **149** of the training set is displayed in the background. (b) CoMSIA STDEVXCOEFF electrostatic contour maps. The most mutagenic molecule**149** of the training set is displayed in the background. (c) CoMSIA STDEVXCOEFF hydrophobic contour maps. The most mutagenic molecule **149** of the training set is displayed in the background. (d) CoMSIA STDEVXCOEFF donor contour maps. The most mutagenic molecule **149** of the training set is displayed in the background. (e) CoMSIA STDEVXCOEFF acceptor contour maps. The most mutagenic molecule **149** of the training set is displayed in the background.

steric, electrostatic and acceptor fields. Fig. 7a and b shows the plot of actual versus predicted log revertants. Looking at those plots one can decipher that the general performance of CoMFA and CoMSIA models for predicting the mutagenic behavior of nitroarenes are almost similar.

4.2. HQSAR

The statistical results of HQSAR analysis is summarized in Table 3. Unlike CoMFA, aligning the molecules is not essential in HQSAR studies as it predicts the activity of compounds

Table 3
Statistical parameters obtained from HQSAR analysis

	Statistical parameters						
Atom count	1–4	2–5	3–6	4–7	5–8	6–9	7–10
r^2 (LOO)	0.666	0.736	0.766	0.739	0.740	0.742	0.726
r^2_{ncv}	0.777	0.930	0.928	0.869	0.865	0.924	0.918
S.E.	0.916	0.534	0.534	0.707	0.714	0.545	0.563
HL	151	353	401	307	401	151	199
N	6	16	13	7	6	10	9

AC, Atom count (fragment size); r^2_{cv} (LOO), cross-validated correlation coefficient; S.E., non-cross-validated standard error; r^2_{ncv}, non-cross-validated correlation coefficient; HL, best hologram length; *N*, optimal number of components.

based on fragment considerations. The best model showed cross validated r^2 of 0.736 with 16 optimal components using leave one out (LOO) method, conventional r^2 of 0.930 and S.E. of 0.534. The hologram length for the best r^2_{cv} was 343. The actual and predicted values for training and test set, obtained from HQSAR analysis are integrated in Tables 6 and 7. The plot of actual and predicted values for the training and test set is shown in Fig. 7c. In HQSAR, the molecules are broken down into smaller structural fragments and these provide critical information about the activity of the compounds. This information may aid in hypothesizing the influence of certain fragments in drugs and drug-like molecules for the mutagenic effect and serve as a valuable tool in screening the molecules in early or late phase of drug discovery. In the present study, the fragments are derived from the aromatic and heteroaromatic nitro compounds and one can trace out their contribution for the mutagenic behaviour of nitro compounds. The color coding for a few molecules representing atom wise contributions is shown in Fig. 5. Colors at the red end of the spectrum reflect unfavorable contributions while colors at the green end reflect favorable contributions to mutagenicity. Atoms with intermediate contributions are colored white. One can decipher from this color coding of which particular atom or group is responsible for the mutagenic behaviour of the molecules and exploit it further for reducing the mutagenicity. For example, replacing the atoms or groups with suitable substitutions in green coded regions may reduce the mutagenicity for that particular class of compounds.

A few fragments with positive and negative contributions to mutagenic activity are shown in Fig. 6. Fragments having positive contributions suggest that they are more mutagenic in comparison with fragments having negative contributions. The analysis of log revertant data and the fragments contribution shows there is a correlation exists between the two. In general, it seems those mutagens showing high log revertant values have positively contributing fragments and those mutagens showing low log revertant values have negatively contributing fragments. The analysis of fragments also highlighted meta substituted fluoro, bromo and chloro fragments (F13, F14, F15 in Fig. 6) showed negative contributions (F13 −0.014, F14 −0.077, F15 −0.019). This agrees with the mutagenic activity results obtained by Shimizu et al. on fluoro and chloro substituted nitro benzenes. They showed that ortho and para substituted analogues were mutagenic where as meta substituted analogues are non-mutagenic [55]. Similar conclusions can be drawn for the fragments with electronegative groups which are positioned meta to nitro group. Considering the electron withdrawing nature of ring nitrogens, a few assumptions were made as follows. Fragments F8 (0.012), F10 (0.151) and F12 (0.145) show positive contributions due to the electron withdrawing tendency of ring nitrogens making N of nitro group or hydroxylamine intermediate more electrophilic, to get attacked by the nucleophile enzyme or DNA bases. Considering the presence of amino group in fragment F9, one can suggest that electron donating tendency of amino to halo substituents will lessen the mutagenic effect of the molecules. On the other hand, the presence of electron withdrawing nitro substituent will enhance the mutagenic effect. The hydroxyl substituent present at ortho to chloro in fragment F6 has more mutagenic effect compared to amino in F9, probably because of its less electron donating effect. It is also seen that nitro attached to non-aromatic unsaturated ring systems (F3) shows less contribution in comparison to the one attached to aromatic ring systems (F5). In this context, the analysis of the whole dataset reveals that saturated ring systems are less mutagenic in comparison to the aromatic ones. A pattern of decline in log revertant values is commonly seen in dataset as one goes from tetrahydro (**92** Table 1) to dihydro (**125** Table 1) to aromatic systems (**103** Table 1).

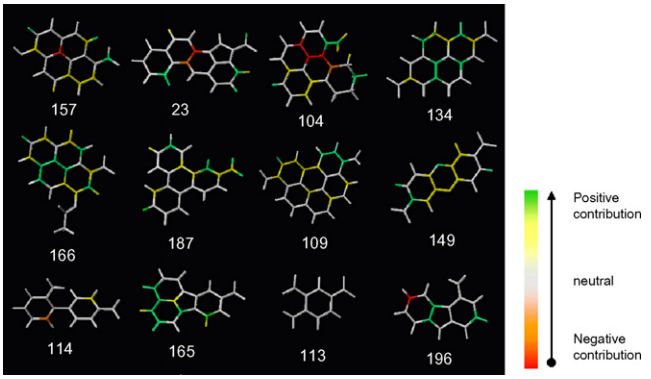

Fig. 5. Contribution map obtained by HQSAR analysis giving atom wise contributions are shown for a few molecules.

4.3. GFA based QSAR

The final model for GFA based QSAR consists of spatial, topological and thermodynamic descriptors. Summary of statistical analysis is shown in Table 4. The actual and predicted values for training and test set, obtained from GFA

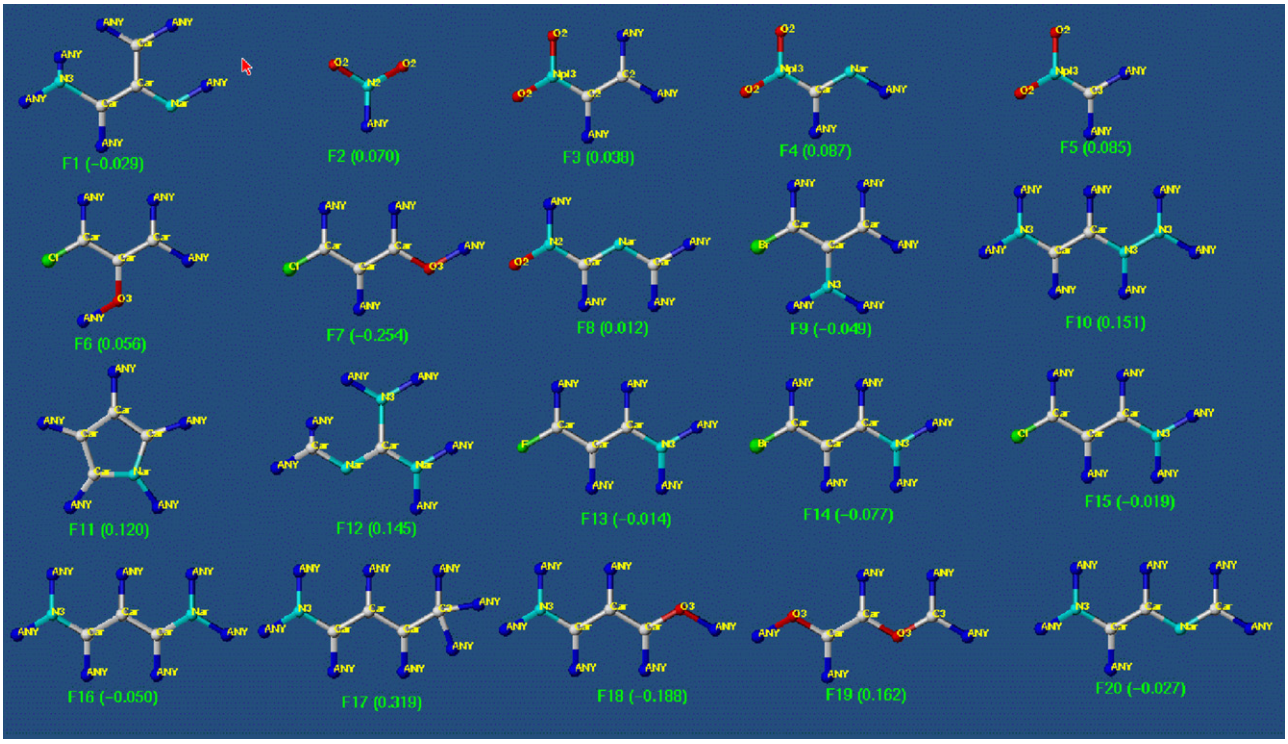

Fig. 6. Few fragments with positive and negative contributions to mutagenicity derived by HQSAR study. Color coding for atoms: red (O), grey (C), green (F, Cl, Br), cyan (N) and blue (any atom).

Table 4
Statistical parameters obtained from GFA analysis

Statistical parameters	
r	0.883
LOF	0.917
r^2	0.779
r^2_{adj}	0.771
F-test	95.824
nNvars	6
r^2_{cv}	0.754
r^2_{bs}	0.779
bsr^2_{er}	0.001

r, Correlation coefficient; LOF, lack of fit score, r^2, squared correlation coefficient, r^2_{adj}, square of adjusted correlation coefficient, F-test, variance related static which compares two models; r^2_{cv}, the square of the correlation coefficient of the cross validation; r^2_{bs}, correlation coefficient of bootstrapping analysis; bsr^2_{er}, error associated with bootstrapping analysis.

analysis are included in Tables 6 and 7. The plot of actual and predicted values for the training and test set is shown in Fig. 7d. Shape of the molecules depicts the role of steric factors in molecules. Table 5 shows the correlation matrix obtained from GFA. One can note that Shadow-nu descriptor which characterizes extents of molecular shadows has good correlation with mutagenicity. The various models developed by us showed the presence of atom type descriptors which characterize the hydrophobicity. Atype_C_25 is one of the atom type AlogP descriptor which was chosen up by genetic function approximation method of variable selection suggesting its crucial role in determining the mutagenicity of compounds. In Atype_C_25, C is part of R- -CR- -R, where R represents any group linked through carbon and - - represents aromatic bonds as in benzene or delocalized bonds in N–O bond in nitro group. Atype_N_76, another AlogP descriptor

Table 5
Correlation matrix obtained from GFA based QSTR analysis

	Correlation matrix					
	kappa-2AM	JX	Atype-C25	Atype-N76	Shadow-nu	log revertant
kappa-2AM	1					
JX	−0.4353	1				
Atype-C25	0.3438	−0.5904	1			
Atype-N76	0.5331	0.2133	−0.0077	1		
Shadow-nu	−0.0508	−0.4106	0.1084	−0.1905	1	
log revertant	0.3676	−0.6445	0.6683	0.2351	0.4557	1

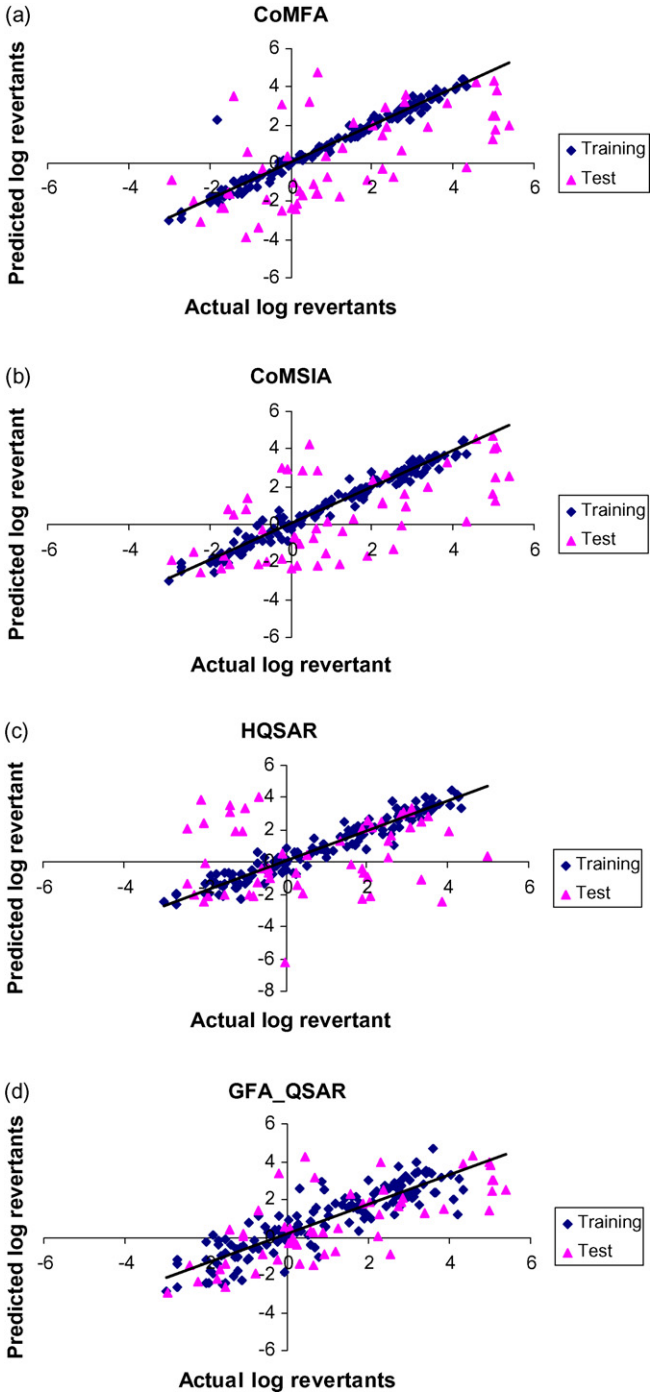

Fig. 7. (a) Plot of actual vs. predicted log revertants of the training and test set molecules by CoMFA. The training set and test set molecules are shown in blue and pink spots, respectively. (b) Plot of actual vs. predicted log revertants of the training and test set molecules by CoMSIA. The training set and test set molecules are shown in blue and pink spots, respectively. (c) Plot of actual vs. predicted log revertants of the training and test set molecules by HQSAR analysis. The training set and test set molecules are shown in blue and pink spots, respectively. (d) Plot of actual vs. predicted log revertants of the training and test set molecules by GFA based QSAR analysis. The training set and test set molecules are shown in blue and pink spots, respectively.

probably has relationship with lipophilicity of either of these Ar-NO_2, R- -N (- -R) - -O′, RO-NO_2. In our case N_76 seems to be associated with Ar-NO_2 as majority of the dataset molecules contain aromatic and heteroaromatic nitro compounds. Ghose and Crippen [56] used the atomic contribution of individual atom types towards the overall hydrophobicity of molecules, where carbon, hydrogen, oxygen, nitrogen, sulfur and halogens are classified into 120 atom types [57]. Halogens and hydrogens were classified by the hybridization and oxidation state of the carbon they are bonded to; carbon atoms were classified by their hybridization state and the chemical nature of their neighboring atoms. The topological descriptor, viz. Balaban index (JX) picked up by GFA plays a role in determining the mutagenicity associated with nitroarenes. The Balaban indices constitute a highly discriminating descriptor whose values do not substantially increase with model size and number of rings present. JX takes into account the relative electronegativity of the carbon in presence of heteroatom. JX showed a negative contribution to the mutagenic behavior in case of nitroarenes. Kier's shape indices are yet another descriptor obtained by GFA runs. These indices compare the model graph with ''minimal'' and ''maximal'' graphs. Kappa 2-AM is second order Kier's alpha modified shape index which take into consideration the contributions of covalent radii and hybridization states. Kappa 2-AM shows a positive contribution to the mutagenicity of nitro compounds.

5. Comparison of QSTR models

The analysis of the QSTR model shows that some classes of nitroarenes are predicted better by 3D QSAR method and some by 2D methods. We studied the predictive ability of different models by analyzing the different classes of nitroarenes using the residual values which are the difference between the actual and predicted log revertants. We used the cut-off as residual >1 for this analysis. Considering the pyrene class of molecules which forms the largest subset, we found CoMFA and CoMSIA predicted these molecules reasonably good. Alignment being one of the important steps in CoMFA and CoMSIA, probably the aligned pyrene analogues were suitable for the right predictions by these methods. Similar predictive trend was observed in biphenyls also. One can notice that fluorene classes of compounds were predicted well by all the models. It was not surprising that 2D methods were better in prediction for certain classes, for example, the nitrobenzene class was predicted well by both HQSAR and GFA methods. Even the indazole classes were predicted very well by GFA in comparison to the 3D methods. The analysis of residuals from Tables 6 and 7 also shows there are many outliers and this was in fact expected as the mutagenicity data used in the study involves data from different laboratories. Hansch and others also mentioned this point in their study that a reliable single laboratory source should derive data, utilizing pure compounds for the study.

In statistical terms, Hansch model sounds more suitable than our present work. Nevertheless, the authors feel, the QSTR models developed in the present study provide enormous information at every stage, compared Hansch study. For example, the steric factors of CoMFA and the acceptor parameters of CoMSIA were not considered implicitly by Hansch models. The fragment wise information from HQSAR analysis is of more value to understand the most and least

Table 6
Training set molecules used for development of different QSTR models

S.no.	Molecules	Actual log revertants	Predicted log revertants			
			CoMFA	CoMSIA	HQSAR	GFA_QSAR
1	**1**	2.11	2.11	2.16	2.02	1.89
2	**2**	−1.24	−1.49	−1.61	−1.57	−0.62
3	**3**	−0.96	−0.84	−0.74	−1.01	−1.13
4	**7**	−0.10	−0.08	−0.87	−0.59	−0.46
5	**8**	1.86	1.74	1.95	1.54	1.37
6	**9**	−1.83	2.27	−1.76	−1.18	0.09
7	**12**	2.43	2.36	2.44	2.48	1.62
8	**13**	3.78	3.71	3.68	3.58	3.14
9	**14**	−1.10	−0.82	−0.67	−0.77	−0.47
10	**15**	2.78	2.62	2.56	2.86	1.66
11	**16**	3.62	3.79	3.50	3.93	2.40
12	**17**	−3.00	−2.98	−2.98	−2.48	−2.85
13	**18**	2.13	1.96	2.31	2.13	2.27
14	**20**	2.51	2.37	2.52	2.76	2.57
15	**21**	0.23	0.42	0.47	0.65	1.16
16	**22**	2.97	3.00	2.78	2.87	2.11
17	**23**	0.86	0.80	1.17	0.76	2.91
18	**25**	1.68	1.68	1.49	2.09	2.12
19	**27**	2.63	2.73	2.79	2.06	2.87
20	**28**	2.68	2.42	2.64	2.17	2.12
21	**29**	3.41	3.57	3.54	3.17	3.37
22	**31**	3.11	3.17	3.13	2.56	2.43
23	**32**	1.54	1.51	1.62	1.96	1.63
24	**34**	−0.30	0.10	0.76	0.02	0.41
25	**36**	0.00	−0.02	−0.30	−0.53	−0.21
26	**37**	0.69	0.68	0.69	0.76	−0.57
27	**38**	−0.30	−0.13	0.39	−0.31	0.32
28	**39**	−1.21	−0.86	−1.26	−1.46	−1.05
29	**41**	3.06	2.95	2.70	2.70	2.14
30	**43**	3.01	3.49	3.41	3.19	3.02
31	**44**	1.99	2.29	2.77	0.81	1.13
32	**46**	0.90	0.93	0.99	0.80	2.54
33	**47**	3.50	3.36	3.32	3.52	3.48
34	**48**	2.95	2.27	2.25	3.29	2.14
35	**49**	2.75	2.96	2.86	3.33	2.48
36	**50**	0.52	0.28	0.60	−0.24	0.11
37	**51**	2.41	2.57	2.42	1.98	2.76
38	**52**	2.41	2.55	2.50	1.74	2.59
39	**53**	3.22	3.44	3.35	3.79	3.48
40	**54**	1.73	1.81	1.89	1.01	1.46
41	**55**	−0.70	−0.52	0.24	−0.09	−0.64
42	**56**	2.06	2.46	2.44	2.65	1.00
43	**57**	0.66	0.57	0.81	0.87	1.33
44	**58**	2.80	2.65	2.90	2.86	2.72
45	**59**	0.26	0.43	0.01	0.43	0.68
46	**60**	0.35	0.26	0.49	0.23	0.31
47	**61**	3.11	3.14	2.69	2.76	2.15
48	**64**	3.19	3.25	3.41	3.02	3.94
49	**65**	−0.05	−0.15	−0.03	−0.03	0.12
50	**66**	−1.52	−1.67	−1.55	−0.65	−0.07
51	**68**	2.82	2.8	2.65	3.00	3.02
52	**69**	3.20	2.67	2.67	3.25	2.43
53	**70**	−0.94	−0.69	−0.88	−1.18	0.74
54	**71**	2.19	2.35	2.18	2.13	2.41
55	**72**	1.43	1.96	2.41	2.49	2.04
56	**73**	−2.00	−2.08	−2.03	−1.31	−1.23
57	**74**	2.76	2.95	2.88	2.61	3.72
58	**75**	1.17	1.33	1.62	0.67	1.77
59	**77**	−1.52	−1.55	−1.44	−1.33	−0.83
60	**78**	−0.23	−0.49	−0.77	−0.21	−0.28
61	**79**	1.79	1.79	1.81	1.29	1.75
62	**82**	3.56	3.59	3.66	3.06	3.38

Table 6 (*Continued*)

S.no.	Molecules	Actual log revertants	Predicted log revertants			
			CoMFA	CoMSIA	HQSAR	GFA_QSAR
63	**83**	1.01	1.01	1.30	0.26	1.52
64	**84**	−1.30	−1.13	−0.36	−0.78	−1.08
65	**85**	0.04	0.08	−0.08	0.12	0.22
66	**86**	0.92	0.74	1.07	1.12	0.65
67	**87**	1.64	1.37	1.59	1.16	1.55
68	**89**	−1.00	−1.25	−1.14	−1.30	−1.20
69	**90**	1.59	1.40	0.94	1.85	1.70
70	**91**	3.08	3.12	3.00	3.09	2.05
71	**92**	1.58	1.75	1.73	1.90	2.12
72	**93**	2.95	2.81	3.01	3.02	2.32
73	**95**	1.51	1.39	1.42	0.95	0.80
74	**97**	2.79	2.75	2.43	2.13	2.32
75	**98**	−1.55	−1.61	−1.51	−1.50	−0.93
76	**99**	1.77	1.74	1.29	1.81	0.31
77	**100**	−1.70	−1.65	−1.52	−0.78	−0.37
78	**102**	2.93	3.08	3.28	3.32	3.25
79	**103**	3.35	2.90	2.99	2.64	2.82
80	**104**	0.70	0.85	0.71	1.49	1.58
81	**105**	0.59	0.48	0.37	0.55	−0.40
82	**106**	1.89	1.92	1.73	1.91	1.26
83	**107**	3.44	3.39	3.34	4.06	3.56
84	**108**	3.67	3.33	2.96	3.36	2.36
85	**109**	3.60	3.88	3.68	3.69	4.69
86	**111**	−1.64	−1.79	−1.99	−1.99	−0.96
87	**112**	1.00	1.08	1.19	0.61	1.14
88	**113**	−1.74	−2.14	−1.88	−1.28	−1.38
89	**114**	−1.70	−1.41	−1.88	−1.31	−0.18
90	**117**	0.03	0.10	−0.08	0.85	0.93
91	**118**	0.30	0.18	0.34	0.62	2.62
92	**119**	−0.19	0.10	0.36	0.32	0.37
93	**120**	−0.82	−1.28	−0.78	−0.83	−0.18
94	**121**	3.16	3.10	3.24	2.88	3.40
95	**122**	1.75	1.59	1.91	1.26	1.51
96	**123**	−0.84	−0.83	−0.05	−0.49	−0.61
97	**125**	3.27	3.02	2.76	3.12	1.73
98	**127**	1.92	1.78	1.95	1.74	1.25
99	**128**	1.91	1.90	1.99	2.23	0.99
100	**129**	−0.07	0.06	−0.25	−0.66	−1.88
101	**130**	−2.70	−2.64	−2.49	−2.65	−2.62
102	**131**	−1.15	−0.95	−1.34	−1.37	−0.52
103	**132**	−1.92	−1.85	−1.72	−2.02	−2.45
104	**133**	−1.40	−0.97	−0.92	−0.62	−0.39
105	**134**	4.25	4.38	4.45	4.06	3.09
106	**135**	−1.21	−1.60	−1.60	−1.57	−1.81
107	**137**	2.62	2.33	2.66	3.74	2.76
108	**138**	−0.41	−0.49	−0.24	−0.63	0.12
109	**139**	−2.00	−1.75	−2.00	−1.96	−2.46
110	**142**	−0.53	−0.49	−0.42	0.25	0.55
111	**144**	−1.54	−1.50	−2.08	−2.07	−2.45
112	**145**	0.25	0.30	0.20	−0.76	0.28
113	**146**	0.19	0.20	0.15	−0.64	0.51
114	**147**	−0.51	−1.06	−0.65	−1.35	−0.37
115	**148**	1.77	1.58	1.49	2.23	1.68
116	**149**	4.34	4.02	3.73	3.33	2.54
117	**150**	−1.61	−1.39	−1.75	−1.43	−2.51
118	**153**	0.58	0.60	0.33	0.44	0.85
119	**154**	0.00	−0.01	−0.07	−0.05	−0.01
120	**155**	−1.89	−1.89	−2.53	−1.90	−1.77
121	**156**	−0.48	−0.78	−0.97	−0.25	0.27
122	**157**	1.34	1.35	1.44	1.53	1.74
123	**158**	2.85	2.85	2.83	3.11	2.04
124	**159**	1.49	1.32	1.52	1.53	1.57
125	**160**	−1.32	−1.58	−1.47	−1.12	−2.10
126	**161**	1.99	2.10	2.06	2.33	3.01

Table 6 (*Continued*)

S.no.	Molecules	Actual log revertants	Predicted log revertants			
			CoMFA	CoMSIA	HQSAR	GFA_QSAR
127	**165**	4.05	3.88	3.61	3.56	3.29
128	**166**	4.22	4.37	4.41	4.22	1.19
129	**168**	−2.70	−2.90	−2.27	−2.00	−1.44
130	**171**	2.66	2.54	2.20	1.07	1.17
131	**173**	2.60	2.58	2.15	1.76	1.84
132	**175**	−0.33	−0.33	−0.53	−0.74	−0.87
133	**176**	0.96	0.57	0.41	0.79	0.53
134	**181**	−1.11	−1.37	−1.04	−2.30	−2.31
135	**182**	−2.70	−2.57	−2.08	−1.94	−1.20
136	**185**	−1.84	−1.79	−2.06	−0.83	−1.95
137	**186**	−0.17	0.16	0.36	0.02	0.28
138	**187**	4.09	4.03	3.72	4.45	2.02
139	**189**	−0.70	−0.58	−0.87	0.61	1.16
140	**191**	0.77	0.77	0.88	0.67	−1.07
141	**192**	−0.22	0.07	0.06	−0.03	1.10
142	**196**	−2.00	−1.91	−2.03	−1.36	−0.23

Table 7
Test set molecules used for development of different QSTR models

S.no.	Molecules	Actual log revertants	Predicted log revertants			
			CoMFA	CoMSIA	HQSAR	GFA_QSAR
1	**4**	2.74	0.69	−0.06	2.19	1.66
2	**5**	−0.79	−3.39	−2.11	−2.11	−1.89
3	**6**	4.99	1.28	1.60	3.87	3.97
4	**10**	2.36	1.92	2.60	2.53	2.49
5	**11**	2.25	1.47	1.16	1.32	0.04
6	**19**	−1.08	0.62	1.37	−0.60	0.17
7	**24**	0.45	3.24	4.24	−6.18	4.29
8	**26**	0.30	−1.65	2.89	−1.39	−1.40
9	**30**	0.67	4.74	2.89	3.07	3.16
10	**33**	2.81	3.20	1.59	3.08	1.82
11	**35**	1.56	2.12	0.28	2.14	2.26
12	**40**	−0.61	−1.87	−1.99	−0.42	−0.93
13	**42**	−1.10	−3.85	0.80	−0.42	0.01
14	**45**	0.15	−2.12	−2.22	−0.65	−0.07
15	**62**	−2.40	−1.94	−1.45	−2.46	−1.50
16	**63**	2.45	1.78	1.45	1.40	2.20
17	**67**	2.26	−0.30	1.10	2.09	1.20
18	**76**	−1.72	−2.34	−2.33	−2.14	−2.25
19	**80**	1.87	−0.88	−1.65	2.04	1.80
20	**81**	0.03	−2.33	−2.35	−0.88	-0.33
21	**88**	−1.54	−1.55	0.78	−2.04	−1.44
22	**94**	3.39	1.89	1.95	2.45	1.28
23	**96**	4.99	2.52	4.70	0.39	1.43
24	**101**	0.78	0.40	2.90	1.34	0.72
25	**110**	−0.07	−1.02	−0.63	0.52	0.53
26	**115**	0.08	−2.51	−1.84	0.42	0.45
27	**116**	−0.23	−2.31	−1.71	−1.92	−1.20
28	**124**	−1.66	1.94	2.33	−2.08	−1.69
29	**126**	2.02	3.11	3.31	2.61	1.84
30	**136**	3.87	−0.71	0.14	1.61	1.49
31	**140**	0.90	−1.59	−2.14	−0.16	−0.89
32	**141**	−1.51	−3.05	−2.58	−2.02	−2.63
33	**143**	−2.22	−2.44	−0.70	−2.44	−2.38
34	**151**	0.11	2.96	2.62	−1.37	−2.88
35	**152**	2.32	3.59	0.92	3.51	3.98
36	**162**	2.85	4.26	4.53	2.85	2.05
37	**163**	4.58	2.50	2.52	2.90	4.36
38	**164**	5.04	1.77	1.21	3.12	2.42
39	**167**	5.06	−1.73	−2.13	3.36	2.99

Table 7 (*Continued*)

S.no.	Molecules	Actual log revertants	Predicted log revertants			
			CoMFA	CoMSIA	HQSAR	GFA_QSAR
40	**169**	1.20	0.37	−1.57	−1.05	−0.74
41	**170**	0.87	−1.62	−2.17	1.93	0.24
42	**172**	0.66	4.32	3.99	−0.67	0.25
43	**174**	5.02	0.82	−0.40	4.04	3.85
44	**177**	1.26	−1.47	−1.01	1.89	0.48
45	**178**	0.23	3.52	0.53	−0.45	−0.37
46	**179**	−1.42	−0.19	0.16	−0.38	0.37
47	**180**	4.33	2.01	2.53	−1.01	3.90
48	**183**	5.39	3.80	4.07	3.36	2.52
49	**184**	5.09	−0.30	−0.32	2.53	3.04
50	**188**	−0.70	1.04	1.23	0.25	0.08
51	**190**	0.57	−1.08	−0.72	−1.24	0.34
52	**193**	−0.22	3.09	2.98	1.90	3.40
53	**194**	0.63	−1.62	−0.22	−2.24	−1.50
54	**195**	−2.94	−0.89	−1.92	−2.01	−2.96
55	**197**	2.54	−0.70	−1.30	−0.03	−0.92

contributing mutagenic fragment patterns which Hansch model failed to bring out. Also different topological descriptors not employed by Hansch model are considered to be important in our GFA analysis.

6. Conclusions

The present work focuses on use of 2D (HQSAR and GFA) and 3D QSAR (CoMFA and CoMSIA) methods to develop predictive models for a structurally diverse set **197** nitroarene molecules. The models developed in the study are statistically significant despite diversity of compound series used for the analysis. The findings from the 3D models shows, to form a stable complex between the mutagen and the DNA, the mutagens must have suitable electronic feature, hydrogen acceptor character, steric bulk and unsaturation of aromatic rings. Fragment based studies provide better insight into the mutagenicity of nitro compounds. The analysis shows the unsaturation of aromatic rings presumably stabilizes the molecule by its resonance effects. The saturated ring systems with dihydro and tetrahydro are of less mutagenic potential in comparison with its unsaturated counterparts due to their lack of resonance stabilization. The other reason that influences the saturated systems is that they are less planar in comparison to their unsaturated counterparts and therefore less likely to form a stable intercalate with DNA. The GFA based model shows thermodynamic and topological descriptors play an important role in characterizing mutagenicity of nitroarenes. Particularly, atomic-level thermodynamic descriptor namely AlogP throws light on hydrophobic features and helps to understand the bilinear model. Fragment fingerprints obtained from HQSAR models also shows the topological importance of the molecules. The color coding from HQSAR highlights which particular atom or groups is responsible for the mutagenic behaviour of the molecules. Overall, the QSTR analyses show there is no single model which can predict well the entire class of nitroarenes as one in the present study. Some classes of nitroarenes are predicted well by 3D QSAR methods and some by the 2D methods.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2007.06.006.

References

[1] T.W. Schultz, M.T.D. Cronin, J.D. Walker, A.O. Aptula, Quantitative structure–activity relationships (QSARS) in toxicology: a historical perspective, Theochem 622 (2003) 1–22.
[2] R. Benigni, Structure–activity relationship studies of chemical mutagens and carcinogens: mechanistic investigations and prediction approaches, Chem. Rev. 105 (2005) 1767–1800.
[3] R.P. Verma, S.B. Mekapati, A. Kurup, C. Hansch, A QSAR review on melanoma toxicity, Bioorg. Med. Chem. 19 (2005) 5508–5526.
[4] S. Khanna, M.E. Sobhia, P.V. Bharatam, Additivity of molecular fields: CoMFA study on dual activators of PPARalpha and PPARgamma, J. Med. Chem. 48 (2005) 3015–3025.
[5] M. Bohm, J. St rzebecher, G. Klebe, Three-dimensional quantitative structure–activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa, J. Med. Chem. 42 (1999) 458–477.
[6] M.L. Hanson, K.R. Solomon, New technique for estimating thresholds of toxicity in ecological risk assessment, Environ. Sci. Technol. 36 (2002) 3257–3264.
[7] M.T.D. Cronin, A.O. Aptula, J.C. Duffy, T.I. Netzeva, P.H. Rowe, I.V. Valkova, T.W. Schultz, Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to Tetrahymena pyriformis, Chemosphere 49 (2002) 1201–1221.
[8] R. Gieleciak, J. Polanski, Modeling robust QSAR. 2. Iterative variable elimination schemes for CoMSA: application for modeling benzoic acid pK_a values, J. Chem. Inf. Model. 47 (2007) 547–556.
[9] L. Tomatis, J. Huff, Evolution of cancer etiology and primary prevention, Environ. Health Perspect. 109 (2001) 458–460.
[10] J. McCann, E. Choi, E. Yamasaki, B.N. Ames, Detection of carcinogens as mutagens in the Salmonella/microsome test: assay of 300 chemicals, Proc. Natl. Acad. Sci. 72 (1975) 5135–5139.
[11] J. McCann, N.E. Spingarn, J. Kobori, B.N. Ames, Detection of carcinogens as mutagens: bacterial tester strains with R factor plasmids, Proc. Natl. Acad. Sci. 72 (1975) 979–983.
[12] E. Zeiger, Carcinogenicity of mutagens: predictive capability of the Salmonella mutagenesis assay for rodent carcinogenicity, Cancer Res. 47 (1987) 1287–1296.

[13] A.M. Richard, International commission for protection against environmental mutagens and carcinogens. Application of SAR methods to non-congeneric data bases associated with carcinogenicity and mutagenicity: issues and approaches, Mutat. Res. 305 (1994) 73–97.
[14] R. Benigni, C. Andreoli, M. Cotta-Ramusino, F. Giorgi, G. Gallo, The electronic properties of carcinogen, and their role in SAR studies of non-congeneric chemicals, Toxicol. Model. 1 (1995) 157–167.
[15] R. Benigni, G. Alessandro, R. Franke, A. Gruska, Quantitative structure–activity relationships of mutagenic and carcinogenic aromatic amines, Chem. Rev. 100 (2000) 3697–3714.
[16] S.C. Basak, D.R. Mills, A.T. Balaban, B.D. Gute, Prediction of mutagenicity of aromatic and heteroaromatic amines from structure: a hierarchical QSAR approach, J. Chem. Inf. Comput. Sci. 41 (2001) 671–678.
[17] P. Gramatica, V. Consonni, M. Pavan, Prediction of aromatic amines mutagenicity from theoretical molecular descriptors, SAR QSAR Environ. Res. 14 (2003) 237–250.
[18] R. Benigni, R. Zito, Designing safer drugs: (Q)SAR-based identification of mutagens and carcinogens, Curr. Top. Med. Chem. 3 (2003) 1289–1300.
[19] D.G. Hoel, J.K. Haseman, M.D. Hogan, J. Huff, E.E. McConnell, The impact of toxicity on carcinogenicity studies: implications for risk assessment, Carcinogenesis 9 (1988) 2045–2052.
[20] M.D. Shelby, The genetic toxicity of human carcinogens and its implications, Mutat. Res. 204 (1988) 3–15.
[21] J. Ashby, R.W. Tennant, Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP, Mutat. Res. 204 (1988) 17–115.
[22] K. Mortelmans, E. Zeiger, The Ames Salmonella/microsome mutagenicity assay, Mutat. Res. 455 (2000) 29–60.
[23] A.T. Maynard, L.G. Pedersen, H.S. Posner, J.D. McKinney, An Ab initio study of the relationship between nitroarene mutagenicity and electron affinity, Mol. Pharmacol. 29 (1986) 629–636.
[24] R. Benigni, in: Y.T. Woo, D.Y. Lai (Eds.), Quantitative Structure–Activity Relationship Models of Mutagens and Carcinogens. Mechanisms of Action of Chemical Carcinogens and Their Role in Structure–Activity Relationships (SAR) Analysis and Risk Assessment, CRC Press, USA, 2003, pp. 41–80.
[25] G. Sabbioni, Hemoglobin binding of nitroarenes and quantitative structure–activity relationships, Chem. Res. Toxicol. 7 (1994) 267–274.
[26] H.S. Rosenkranz, E.C. McCoy, R. Mermelstein, W.T. Speck, A cautionary note the use of nitroreductase-deficient strains of *Salmonella typhimurium* for the detection of nitroarenes as mutagens in complex mixtures including diesel exhausts, Mutat. Res. 91 (1981) 103–105.
[27] T.C. Pederson, J.S. Siak, The role of nitroaromatic compounds in the direct-acting mutagenicity of diesel particle extracts, J. Appl. Toxicol. 1 (1981) 54–60.
[28] R. Miri, H. Niknahad, G. Vesal, A. Shafiee, Synthesis and calcium channel antagonist activities of 3-nitrooxyalkyl, 5-alkyl 1,4-dihydro-2,6-dimethyl-4-(1-methyl-5-nitro-2-imidazolyl)-3,5-pyridinedicarboxylates, Farmaco 57 (2002) 123–128.
[29] A.K. Debnath, C. Hansch, K.H. Kim, Y.C. Martin, Mechanistic interpretation of the genotoxicity of nitrofurans (antibacterial agents) using quantitative structure–activity relationships and comparative molecular field analysis, J. Med. Chem. 36 (1993) 1007–1016.
[30] H. Tokiwa, Y. Ohnishi, Mutagenicity and carcinogenicity of nitroarenes and their sources in the environment, Crit. Rev. Toxicol. 17 (1986) 23–60.
[31] J.M. Goldring, L.M. Ball, R. Sangaiah, A. Gold, Mutagenic activity of nitro-substituted cyclopenta-fused polycyclic aromatic hydrocarbons towards *Salmonella typhimurium*, Mutat. Res. 187 (1987) 67–77.
[32] H.S. Rosenkranz, R. Mermelstein, Mutagenicity and genotoxicity of nitroarenes. All nitro-containing chemicals were not created equal, Mutat. Res. 114 (1983) 217–267.
[33] B. Zielinska, J. Arey, W.P. Harger, R.W. Lee, Mutagenic activities of selected nitrofluoranthene derivatives in *Salmonella typhimurium* strains TA98, TA98NR and TA98/1,8-DNP6, Mutat. Res. 206 (1988) 131–140.
[34] A.K. Debnath, R.L. Lopez de Compadre, G. Debnath, A.J. Shusterman, C. Hansch, Structure–activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. Correlation with molecular orbital energies and hydrophobicity, J. Med. Chem. 34 (1991) 786–797.
[35] SYBYL Molecular Modeling System, version 7.1; Tripos, Inc., St. Louis, MO, 63144-2913.
[36] Cerius2 version 4.10, Accelrys Inc., 6985 Scranton Road, San Diego, CA, USA.
[37] M.J.D. Powell, Restart procedures for the conjugate gradient method, Math. Program. 12 (1977) 241–254.
[38] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electro negativity—a rapid access to atomic charges, Tetrahedron 36 (1980) 3219–3228.
[39] M.T. Makhija, V.M. Kulkarni, 3D-QSAR and molecular modeling of HIV-1 integrase inhibitors, J. Comput. Aided Mol. Des. 16 (2002) 181–200.
[40] A.V. Raichurkar, V.M. Kulkarni, Understanding the antitumor activity of novel hydroxysemicarbazide derivatives as ribonucleotide reductase inhibitors using CoMFA and CoMSIA, J. Med. Chem. 46 (2003) 4419–4427.
[41] R.D. Cramer, D.E. Patterson, J.D. Bunce, Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[42] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, J. Med. Chem. 37 (1994) 4130–4136.
[43] V.N. Viswanadhan, A.K. Ghose, G.R. Revenkar, R.J. Robins, Chem. Inf. Comput. Sci. 28 (1989) 163.
[44] G. Klebe, The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands, J. Mol. Biol. 237 (1994) 212–235.
[45] C. de Gregorio, L.B. Kier, L.H. Hall, QSAR modeling with the electrotopological state indices: corticosteroids, J. Comput. Aided Mol. Des. 12 (1998) 557–561.
[46] L.B. Kier, L.H. Hall, An electrotopological-state index for atoms in molecules, Pharm. Res. 7 (1990) 801–807.
[47] E. Estrada, E. Uriarte, Recent advances on the role of topological indices in drug discovery research, Curr. Med. Chem. 8 (2001) 1573–1588.
[48] D. Rogers, A.J. Hopfinger, Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships, J. Chem. Inf. Comput. Sci. 34 (1994) 854–866.
[49] L.M. Shi, F. Yi, T.G. Myers, P.M. O'Connor, K.D. Paull, S.H. Friend, J.N. Weinstein, Mining the NCI anticancer drug discovery databases: genetic function approximation for the QSAR study of anticancer ellipticine analogues, J. Chem. Inf. Comput. Sci. 38 (1998) 189–199.
[50] C. Glende, M. Klein, H. Schmitt, L. Erdinger, G. Boche, Transformation of mutagenic aromatic amines into non-mutagenic species by alkyl substituents. Part II: alkylation far away from the amino function, Mutat. Res. 515 (2002) 15–38.
[51] M. Klein, U. Voigtmann, T. Haack, L. Erdinger, G. Boche, From mutagenic to non-mutagenic nitroarenes: effect of bulky alkyl substituents on the mutagenic activity of 4-nitrobiphenyl in Salmonella typhimurium—part I. Substituents ortho to the nitro group and in 2′-position, Mutat. Res. 467 (2000) 55–68.
[52] M. Nohara, T. Hirayama, Y. Fujioka, S. Ozasa, E. Ibuki, S. Fukui, Relationship between mutagenic potency in *Salmonella typhimurium* and chemical structure of amino- and nitro-substituted biphenyls, Mutat. Res. 149 (1985) 9–15.
[53] A.K. Debnath, R.L. Lopez de Compadre, A.J. Shusterman, C. Hansch, Quantitative structure–activity relationship investigation of the role of hydrophobicity in regulating mutagenicity in the Ames test: 2. Mutagenicity of aromatic and heteroaromatic nitro compounds in *Salmonella typhimurium* TA100, Environ. Mol. Mutagen. 19 (1992) 53–70.
[54] A.K. Debnath, A.J. Shusterman, R.L. Lopez de Compadre, C. Hansch, International Commission for Protection Against Environmental Mutagens and Carcinogens. The importance of the hydrophobic interaction in the mutagenicity of organic compounds, Mutat. Res. 305 (1994) 63–72.
[55] M. Shimizu, Y. Yasui, N. Matsumoto, Structural specificity of aromatic compounds with special reference to mutagenic activity in Salmonella

typhimurium—a series of chloro- or fluoro-nitrobenzene derivatives, Mutat. Res. 116 (1983) 217–238.

[56] A.K. Ghose, G.M. Crippen, Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure–activity relationships. 2. Modeling dispersive and hydrophobic interactions, J. Chem. Inf. Comput. Sci. 27 (1987) 21–35.

[57] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, R.K. Robins, Atomic physicochemical parameters for three dimensional structure directed quantitative structure–activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics, J. Chem. Inf. Comput. Sci. 29 (1989) 163–172.