Calculating Chemical Oxidizabilities of Organic Compounds from Molecular Structure by QSAR Techniques

Takahiro SUZUKI *

Abstract

The method of partial least squares (PLS) and artificial neural network modeling employing back-propagation learning strategy (ANN) have been applied to predict chemical oxidizabilities (COD) of organic compounds. The experimental COD data based on three different oxidation methods, the measurements using permanganate in acid solution (COD_{KMnO_4}) and dichromate/H_2SO_4 without ($COD_{K_2Cr_2O_7}$) and with Ag^+ as a catalyst ($COD_{K_2Cr_2O_7\text{-}Ag^+}$), have been collected from the literature for 390 organic compounds containing C, H, O, N, S, F, Cl, and Br atoms. Two sets of molecular descriptors to characterize the chemicals, physico-chemical descriptors and structural ones, were employed. The PLS approach gave statistically significant models for COD of some homologous series of the compounds. For the ANN approach, two separate models with one hidden layer, one based on the physicochemical parameters and the second based on molecular structural features as the inputs into the networks, were developed. The analysis includes different partitionings of the data set into training and prediction sets, and different number of hidden-layer neurons of the neural networks. It was confirmed that both ANN models can be applied to structurally diverse organic compounds. The ANN approach, which has an inherent ability to provide non-linear and cross product terms, gave global models for the different COD data employed.

Key words: biodegradation, chemicals, COD, neural networks, PLS

1. Introduction

In the generation and discovery of new synthetic organic chemicals, there is a need to know many physicochemical properties together with their target properties (Dearden 1994). Of all the properties required for the stages in molecular design and development, potential oxidization or aerobic biodegradation are important from the viewpoint of *in vivo* biotransformation or metabolism. If a chemical is not mineralized or broken down to nontoxic products in the environment, it can impact an ecosystem at several points, causing

* Natural Science Laboratory, Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, Japan

health, economic, and aesthetic problems. Biodegradation is one of the most important breakdown mechanism that takes place in aquatic and terrestrial ecosystems, due to the natural occurrence of many populations of microorganisms capable of breaking down organic chemicals. The eventual mineralization of organic compounds can be attributed almost entirely to biodegradation. Among many descriptors of biodegradation, there has been several works on the quantitative correlations of various physicochemical and/or structural parameters with biochemical oxygen demand (BOD) of organic compounds (Dearden & Nicholson 1986; Niemi et al 1987; Babeu & Vaishnav 1987; Howard et al. 1992; Zakarya et al 1993).

Several schemes for predicting biodegradability of chemicals have been based on the ratio between BOD and either COD (chemical oxygen demand) or UOD (ultimate oxygen demand). A number of chemicals have been classified into categories of biodegradability on the basis of the BOD/COD ratio (Lyman et al 1974). According to the BOD_5/COD ratios for various organic compounds, compounds with the ratio less than 0.01 were classified as relatively undegradable, between 0.01 and 0.1 as moderately degradable, and greater than 0.1 as relatively degradable. A standard test to compare the biodegradability of organic compounds was developed based on measurement of COD decrease (Pitter 1976). The decrease in the COD of a substance was evaluated until no further decrease was observed using activated sludge inoculum with 20 days of assimilation to the substrate. The percent decrease of total COD was calculated as well as the degradation rate, in units of mg COD removed per gram of initial biomass (dry weight) of inoculum per hour. Although these works have had some limited success as a result of the limited amount of data, the quantitative structure-degradation relationships for chemical oxygen demand (COD) are still poorly understood.

The concepts of BOD and COD assume that all carbon is assimilated into new biomass and that the transformation from substrate to biomass is not inhibited by the compound under investigation or by any other substances in the test medium. COD and BOD may involve different sites of reaction and degrees of reactivity. Moreover, the reactivity of a site in COD depends on the reagent used. However, numerous rules of thumb qualitatively relate biodegradation to structural features such as branching, chain length, functional groups, halogens, and saturation (Scow 1982) would be applicable to the case of COD (e.g., adding halogens, saturation, branching or short chains increases persistence for both aerobic biodegradation and chemical oxidation; adding functional groups such as -OH, $-CO_2H$, ester, aldehyde, amino, nitro, -S-, or thiol increase both aerobiodegrability and chemical oxidizability). The typical rules of thumb for COD are listed in Table 1. These generalizations are applicable only to the specific groups of chemicals in which they have been observed.

In our preliminary study, the applicability of two relatively new modeling tools, the method of partial least squares (PLS) (Wold et al 1984) and artificial neural networks (ANNs)

Table 1. Typical rules of thumb for chemical oxidizabilities (COD) of organic compounds.

- Unsaturated aliphatics are more readily oxidized than corresponding saturated hydrocarbons.
- Alcohols, ethers, aldehydes, acids, esters, amines, nitros, sulfides, and thiols et al., are more susceptible to oxidation than the corresponding non-substituted compounds.
- Halogenated hydrocarbons are less degradable than corresponding organic molecules having other functional groups.
- In the homologous series, the molecules having two carbon atoms are more resistant to oxidation by the method using dichromate/H_2SO_4 without Ag^+.
- The permanganate-COD mostly reaches only up to 10%, often equals 0% of the theoretical oxidizability values.
- The dichromate-COD with Ag^+ gives nearly 100% of the theory for most degradable compounds.

(Zupan & Gasteiger 1993), to the COD problem and got a promising result (Suzuki et al 1997). The present study is an attempt to develop a prediction scheme for the COD values for various organic compounds by employing the ANNs with compared to that by PLS. The successful applications of the ANNs to correlate the kinetics of biodegradation have been reported (Zitko 1991; Tabak & Govind 1993).

2. Materials and Methods

Data sets

The experimental COD data based on three different oxidation methods, the measurements using permanganate in acid solution (COD_{KMnO_4}) and dichromate/H_2SO_4 without ($COD_{K_2Cr_2O_7}$) and with Ag^+ as a catalyst ($COD_{K_2Cr_2O_7\text{-}Ag^+}$), were taken from a report (Janicke 1983). The experimental COD values of 390 diverse organic compounds containing C, H, O, N, S, F, Cl, and Br atoms are conventionally expressed in the dimensionless form as the percentage of theoretical values:

$$\text{COD (\%)} = (\text{measured COD}/\text{theoretical COD}) \times 100 \qquad (1)$$

It has been known that reaching oxidizabilities with most organic substances of often nearly 100% of the theory, the dichromate-COD with Ag^+ proves sufficient and adequate for the ingredients of industrial effluents generally dealt with. In contrast, the permanganate-COD mostly reaches only up to 10%, often equals 0%, of the theory. The COD data for the entire data set of 390 organic chemicals analyzed in this study is listed in Table 2 together with the SMILES (Simplified Molecular Input Line Entry System) (Weininger & Weininger 1990).

Table 2. The databank of COD values for organic compounds examined in this study.

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
	〈aliphatic hydrocarbons〉						
1	isoprene	78-79-5	C5H8	C=CC(C)=C	96	100	3
2	cyclohexene	110-83-8	C6H10	C1CCC=CC1	75	83	3
3	cyclohexane	110-82-7	C6H12	C1CCCCC1	3	5	0
4	*n*-hexane	110-54-3	C6H14	CCCCCC	1	2	0
5	1-heptyne	628-71-7	C7H12	C#CCCCCC	48	64.5	3
6	1-heptene	592-76-7	C7H14	C=CCCCCC	38	73	0
7	cycloheptane	291-64-5	C7H14	C1CCCCCC1	7	9.5	0
8	methylcyclohexane	108-87-2	C7H14	CC1CCCCC1	2	2	0
9	*n*-heptane	142-82-5	C7H16	CCCCCCC	1	2	0
10	2,2,4-trimethylpentane	540-84-1	C8H18	CC(C)(C)CC(C)C	1	1	0
11	limonene	5989-27-5	C10H16	C1C(C)=CCC(C(C)=C)C1	76	93	3
12	decahydronaphthalene	493-01-6	C10H18	C12CCCCC2CCCC1	7	7	0
13	*n*-decane	124-18-5	C10H22	CCCCCCCCCC	1	1	0
14	*n*-hexadecane	544-76-3	C16H34	CCCCCCCCCCCCCCCC	1.5	1	0
	〈aromatic hydrocarbons〉						
15	benzene	71-43-2	C6H6	c1ccccc1	63	65	0
16	methylbenzene	108-88-3	C7H8	c1ccccc1C	39	40	0
17	styrene	100-42-5	C8H8	c1ccccc1C=C	64	74	34
18	1,2-dimethylbenzene	95-47-6	C8H10	c1c(C)c(C)ccc1	51.5	72.5	1
19	1,3-dimethylbenzene	108-38-3	C8H10	c1c(C)cc(C)cc1	45	58	0
20	1,4-dimethylbenzene	106-42-3	C8H10	c1c(C)ccc(C)c1	51	50	0
21	ethylbenzene	100-41-4	C8H10	c1ccc(CC)cc1	66	75	0
22	isopropylbenzene	98-82-8	C9H12	c1ccccc1C(C)C	73	76	0
23	indan	496-11-7	C9H10	c1cc2CCCc2cc1	78	79	1
24	napththalene	91-20-3	C10H8	c1cc2ccccc2cc1	100	100	8
25	1,2,3,4-tetrahydronaphthalene	119-64-2	C10H12	c12CCCCc2cccc1	85	90.5	4
26	1,2-diethylbenzene	135-01-3	C10H14	CCc1ccccc1CC	72	89	0

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
27	*tert*-butylbenzene	98-06-6	C10H14	c1ccccc1C(C)(C)C	54	77	0
28	1,2,4,5-tetramethylbenzene	95-93-2	C10H14	c1(C)c(C)cc(C)c(C)c1	66	76	1.5
29	1-methylnaphthalene	90-12-0	C11H10	c1cc2cccc(C)c2cc1	100	100	7
30	acenaphthylene	208-96-8	C12H8	c1c2C=Cc(c23)cccc3cc1	95	99	26
31	biphenyl	92-52-4	C12H10	c1ccccc1c2ccccc2	84	91	0
32	acenaphthene	83-32-9	C12H10	c1c2CCc(c23)cccc3cc1	100	100	17
33	*n*-hexylbenzene	1077-16-3	C12H18	CCCCCCc1ccccc1	16	25	0
34	fluorene	86-73-7	C13H10	c1cc2c3ccccc3Cc2cc1	45	57	6
35	diphenylmethane	101-81-5	C13H12	c1ccccc1Cc2ccccc2	75	76	0
36	anthracene	120-12-7	C14H10	c1cc2cc3ccccc3cc2cc1	62	68	1
37	phenanthrene	85-01-8	C14H10	c1cc2c3ccccc3ccc2cc1	48	63.5	1
38	*trans*-stilbene	103-30-0	C14H12	c1ccccc1C=Cc2ccccc2	100	100	2
39	guaiazulene	489-84-9	C15H18	C1=C2C(C)=CC=C(C(C)C)C=C2C(C)=C1	80	98	27
40	fluoranthene	206-44-0	C16H10	c1c(c34)c2ccccc2c3cccc4cc1	88	92	5
41	chrysene	218-01-9	C18H12	c1cc2ccc3c4ccccc4ccc3c2cc1	69.5	70	0
42	1,4-diphenylbenzene	92-94-4	C18H14	c1ccccc1c2ccc(c3ccccc3)cc2	83	79.5	0.5
43	*n*-dodecylbenzene	123-01-3	C18H30	c1ccccc1CCCCCCCCCCCC	17	34	0
	〈alcohols〉						
44	methanol	67-56-1	CH4O	CO	100	99	2
45	ethanol	64-17-5	C2H6O	CCO	40	99	3
46	2-propyn-1-ol	107-19-7	C3H4O	C#CCO	86.5	96	51
47	1-propanol	71-23-8	C3H8O	CCCO	73.5	99	2
48	2-propanol	67-63-0	C3H8O	CC(O)C	72	90	1
49	3-butyn-2-ol	65337-13-5	C4H6O	C#CC(O)C	66	100	28
50	1-butanol	71-36-3	C4H10O	CCCCO	77.5	100	2
51	2-methyl-2-propanol	75-65-0	C4H10O	CC(C)(C)O	76	98	3.5
52	cyclopentanol	96-41-3	C5H10O	C1CC(O)CC1	76	96.5	2.5
53	1-pentanol	71-41-0	C5H12O	CCCCCO	81	99.5	1
54	2-methyl-2-butanol	75-85-4	C5H12O	CCC(O)(C)C	58	99	9.5
55	cyclohexanol	108-93-0	C6H12O	C1CCC(O)CC1	81	100	3

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
56	*cis*-3-hexen-1-ol	928-96-1	C6H12O	CCC=CCCO	84	100	27
57	1-hexanol	111-27-3	C6H14O	CCCCCCO	77	96.5	1
58	2-hexanol	626-93-7	C6H14O	CC(O)CCCC	74	100	1
59	3-hexanol	623-37-0	C6H14O	CCC(O)CCC	68	100	1
60	benzyl alcohol	100-51-6	C7H8O	c1ccccc1CO	90	98	7
61	1-octanol	111-87-5	C8H18O	CCCCCCCCO	70	98	1
62	(-)menthol	2216-51-5	C10H20O	OC1CC(C)CCC1C(C)C	78	91	1
63	1-decanol	112-30-1	C10H22O	CCCCCCCCCCO	77.5	84	0
64	1-octadecanol	112-92-5	C18H38O	CCCCCCCCCCCCCCCCCCO	86	92	0
	〈phenols〉						
65	phenol	108-95-2	C6H6O	c1ccccc1O	98	99	95
66	2-methylphenol	95-48-7	C7H8O	Cc1ccccc1O	84	100	68.5
67	3-methylphenol	108-39-4	C7H8O	c1ccc(O)cc1C	85	100	65
68	4-methylphenol	106-44-5	C7H8O	c1cc(O)ccc1C	86	96	68
69	3,4-dimethylphenol	95-65-8	C8H10O	c1c(C)c(C)cc(O)c1	75	97	45
70	2,5-dimethylphenol	95-87-4	C8H10O	c1c(C)c(O)cc(C)c1	73	99	53
71	2,6-dimethylphenol	576-26-1	C8H10O	c1c(C)c(O)c(C)cc1	73	100	49
72	4-*n*-propylphenol	645-56-7	C9H12O	c1cc(O)ccc1CCC	89	100	50
73	4-isopropylphenol	99-89-8	C9H12O	c1cc(O)ccc1C(C)C	91	100	50
74	2,3,5-trimethylphenol	697-82-5	C9H12O	c1c(O)c(C)c(C)cc(C)1	73	100	42
75	1-naphthol	90-15-3	C10H8O	c1cc2c(O)cccc2cc1	97	99	45
76	2-naphthol	135-19-3	C10H8O	c1cc2cc(O)ccc2cc1	99	100	56
77	4-*tert*-butylphenol	98-54-4	C10H14O	c1cc(O)ccc1C(C)(C)C	87	99	36
78	5-isopropyl-2-methylphenol	499-75-2	C10H14O	c1(C)c(O)cc(C(C)C)cc1	75	95	38
79	5-methyl-2-isopropylphenol	89-83-8	C10H14O	c1(C(C)C)c(O)cc(C)cc1	78	98.5	41
80	2-*tert*-butyl-4-methylphenol	2409-55-4	C11H16O	c1c(O)c(C(C)(C)C)cc(C)c1	77	98	35
81	2-phenylphenol	90-43-7	C12H10O	c1c(O)c(c2ccccc2)cc1	99	99	52
82	2,6-di-*tert*-butylphenol	128-39-2	C14H22O	c1c(C(C)(C)C)c(O)c(C(C)(C)C)cc1	50	57	1.5
83	*p*-nonylphenol	104-40-5	C15H24O	c1cc(O)ccc1CCCCCCCCC	76	94	11.5
84	2,6-di-*tert*-butyl-4methylphenol	128-37-0	C15H24O	c1c(C(C)(C)C)c(O)c(C(C(C)(C)C)cc1C	72	94	1

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
	〈ethers/O-containing ring〉						
85	furan	110-00-9	C4H4O	O1C=CC=C1	100	100	44.5
86	tetrahydrofuran	109-99-9	C4H8O	C1COCC1	86	100	2
87	diethyl ether	60-29-7	C4H10O	CCOCC	32	81	0.5
88	2-methylfuran	534-22-5	C5H6O	C1=C(C)OC=C1	100	100	22.5
89	isopropyl ether	108-20-3	C6H14O	CC(C)OC(C)C	68	92	0.5
90	methoxybenzene	100-66-3	C7H8O	c1ccccc1OC	91	94	23
91	benzofuran	271-89-6	C8H6O	c1cc2C=COc2cc1	86	87	66
92	*n*-butyl ether	142-96-1	C8H18O	CCCCOCCCC	70	84.5	0
93	dibenzofuran	132-64-9	C12H8O	c1cc2c3ccccc3Oc2cc1	76	76	3
94	diphenyl ether	101-84-8	C12H10O	c1ccccc1Oc2ccccc2	93	99	1.5
95	2-ethoxynaphthalene	93-18-5	C12H12O	c1cc2cc(OCC)ccc2cc1	83	100	22
96	benzylether	103-50-4	C14H14O	c1ccccc1COCc2ccccc2	79	92	2
	〈aldehydes〉						
97	2-propenal	107-02-8	C3H4O	C=CC=O	100	100	39
98	crotonaldehyde	123-73-9	C4H6O	CC=CC=O	65	96.5	54.5
99	*n*-butanal	123-72-8	C4H8O	CCCC=O	78	98	5
100	benzaldehyde	100-52-7	C7H6O	c1ccccc1C=O	100	100	15
101	*n*-decyl aldehyde	112-31-2	C10H20O	CCCCCCCCCC=O	74	90	0.5
	〈ketones〉						
102	2-propanone	67-64-1	C3H6O	CC(=O)C	62.5	91	0
103	2-butanone	78-93-3	C4H8O	CC(=O)CC	45	100	0.5
104	cyclohexanone	108-94-1	C6H10O	C1C(=O)CCCC1	83.5	98.5	7.5
105	4-methyl-2-pentanone	108-10-1	C6H12O	CC(=O)CC(C)C	65.5	95	0
106	2-methylcyclohexanone	583-60-8	C7H12O	C1C(C)CCC(=O)C1	75	100	4
107	acetophenone	98-86-2	C8H8O	c1ccc(C(=O)C)cc1	97	97.5	1
108	2-octanone	111-13-7	C8H16O	CC(=O)CCCCCC	61	91	0
109	1-indanone	83-33-0	C9H8O	c1c2CCC(=O)c2ccc1	98	99	4

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
110	1-phenyl-1-propanone	93-55-0	C9H10O	c1ccccc1C(=O)CC	83.5	98	1.5
111	isophorone	78-59-1	C9H14O	C1C(C)(C)CC(C)=CC(=O)1	77	95	26
112	1-phenyl-1-butanone	495-40-9	C10H12O	c1ccccc1C(=O)CCC	87.5	98.5	1
113	benzophenone	119-61-9	C13H10O	c1ccccc1C(=O)c2ccccc2	99	100	0
114	anthrone	90-44-8	C14H10O	c1c2C(=O)c3ccccc3Cc2ccc1	43.5	57	1
	〈acids〉						
115	acetic acid	64-19-7	C2H4O2	CC(=O)O	6	100	0
116	acrylic acid	79-10-7	C3H4O2	C=CC(=O)O	96	95	73
117	propanoic acid	79-09-4	C3H6O2	CCC(=O)O	71	97.5	0
118	crotonic acid	3724-65-0	C4H6O2	CC=CC(=O)O	60.5	99	58
119	vinylacetic acid	625-38-7	C4H6O2	C=C(C)C(=O)O	94	97	59
120	butanoic acid	107-92-6	C4H8O2	CCCC(=O)O	74	97	0
121	2-methylpropanoic acid	79-31-2	C4H8O2	CC(C)C(=O)O	73.5	100	0
122	pentanoic acid	109-52-4	C5H10O2	CCCCC(=O)O	73	96	0
123	pivalic acid	75-98-9	C5H10O2	CC(C)(C)C(=O)O	68	97	0
124	(*E*,*E*)-2,4-hexadienoic acid	110-44-1	C6H8O2	CC=CC=C(=O)O	79.5	97	57
125	hexanoic acid	142-62-1	C6H12O2	CCCCCC(=O)O	71	100	0
126	benzoic acid	65-85-0	C7H6O2	c1ccccc1C(=O)O	100	100	1
127	octanoic acid	124-07-2	C8H16O2	CCCCCCCC(=O)O	75	100	0.5
128	decanoic acid	334-48-5	C10H20O2	CCCCCCCCCC(=O)O	77	96	1
129	hexadecanoic acid	57-10-3	C16H32O2	CCCCCCCCCCCCCCCC(=O)O	35	39	1
130	9,12,15-octadecatrienoic acid	463-40-1	C18H30O2	CCC=CCC=CCC=CCCCCCCCC(=O)O	85	96	36
131	9,12-octadecadienoic acid	60-33-3	C18H32O2	CCCCCC=CCC=CCCCCCCCC(=O)O	89	100	26
132	9-octadecenoic acid	2027-47-6	C18H34O2	CCCCCCCCC=CCCCCCCCC(=O)O	82	97.5	10
133	octadecanoic acid	57-11-4	C18H36O2	CCCCCCCCCCCCCCCCCC(=O)O	50	59.5	0.5
134	docosanoic acid	112-85-6	C22H44O2	CCCCCCCCCCCCCCCCCCCCCC(=O)O	87	99	0
	〈esters〉						
135	methyl acetate	79-20-9	C3H6O2	CC(=O)OC	49	100	1
136	vinyl acetate	108-05-4	C4H6O2	CC(=O)OC=C	24	90	18

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
137	methyl acrylate	96-33-3'	C4H6O2	C=CC(=O)OC	99	100	40.5
138	ethyl acetate	141-78-6	C4H8O2	CC(=O)OCC	20.5	84	1
139	methyl propionate	554-12-1	C4H8O2	CCC(=O)OC	72	97.5	0.5
140	ethyl acrylate	140-88-5	C5H8O2	C=CC(=O)OCC	64	89	31
141	methylmethacrylate	80-62-6	C5H8O2	C=C(C)C(=O)OC	85	100	30
142	butyl acetate	123-86-4	C6H12O2	CC(=O)OCCCC	72.5	100	0.5
143	isobutyl acetate	110-19-0	C6H12O2	CC(=O)OCC(C)C	63	100	1
144	*tert*-butyl acetate	540-88-5	C6H12O2	CC(=O)OC(C)(C)C	59	82	3
145	ethyl butyrate	105-54-4	C6H12O2	CCCC(=O)OCC	61	95	1
146	butyl acrylate	141-32-2	C7H12O2	C=CC(=O)OCCCC	83	96	21
147	phenyl acetate	122-79-2	C8H8O2	c1ccccc1OC(=O)C	79.5	100	38
148	cyclohexyl acetate	622-45-7	C8H14O2	CC(=O)OC1CCCCC1	72	98	1
149	benzyl benzoate	120-51-4	C14H12O2	c1ccccc1C(=O)OCc2ccccc2	98	95.5	0
	〈amines/imines/pyridines〉						
150	ethyleneimine	151-56-4	C2H5N	C1NC1	86	88	2
151	dimethylamine	124-40-3	C2H7N	CNC	2	2	0.5
152	ethylamine	75-04-7	C2H7N	CCN	18	27	0
153	allylamine	107-11-9	C3H7N	C=CCN	98	100	52
154	trimethylamine	75-50-3	C3H9N	CN(C)C	1	1	1
155	1*H*-pyrrole	109-97-7	C4H5N	C1=CC=CN1	89	95	83.5
156	pyrrolidine	123-75-1	C4H9N	C1CNCC1	92.5	94.5	1
157	diethylamine	109-89-7	C4H11N	CCNCC	12	15	0.5
158	*n*-butylamine	109-73-9	C4H11N	CCCCN	75	97	0.5
159	pyridine	110-86-1	C5H5N	n1ccccc1	0	1	0.5
160	piperidine	110-89-4	C5H11N	C1CCCNC1	68.5	91	1
161	aniline	62-53-3	C6H7N	c1ccccc1N	96	98	87.5
162	triethylamine	121-44-8	C6H15N	CCN(CC)CC	8	10	1
163	*n*-hexylamine	111-26-2	C6H15N	CCCCCCN	62	91	1
164	cyclohexylamine	108-91-8	C6H13N	C1CCCCC1N	82	94	0
165	2-methylaniline	95-53-4	C7H9N	Nc1ccccc1C	82	100	61.5

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
166	*N*-methylaniline	100-61-8	C7H9N	c1ccccc1NC	92	97.5	51
167	1*H*-indole	120-72-9	C8H7N	C1=Cc2ccccc2N1	96	96	89.3
168	2,4-dimethylaniline	95-68-1	C8H11N	c1c(N)c(C)cc(C)c1	81	100	57
169	*N*-ethylaniline	103-69-5	C8H11N	c1ccccc1NCC	76	88	71
170	*N*,*N*-dimethylaniline	95-68-1	C8H11N	c1ccccc1N(C)C	74	82	66
171	2,4,6-trimethylpyridine	108-75-8	C8H11N	n1c(C)cc(C)cc1C	46.5	75.5	0
172	quinoline	91-22-5	C9H7N	c1cc2ncccc2cc1	85	91.8	7
173	2-methyl-6-ethylaniline	24559-06-2	C9H13N	c1c(C)c(N)c(CC)cc1	79.5	100	47
174	*N*-ethyl-m-toluidine	102-27-2	C9H13N	c1c(NCC)cc(C)cc1	71	84	66.5
175	*N*,*N*-diethylaniline	91-66-7	C10H15N	c1ccccc1N(CC)CC	60	78	51
176	tri-*n*-butylamine	102-82-9	C12H27N	CCCCN(CCCC)CCCC	73	94	0.5
177	benzo[*h*]quinoline	230-27-3	C13H9N	c1cc2c3ncccc3ccc2cc1	82	93	17.5
178	acridin	260-94-6	C13H9N	c1cc2nc3ccccc3cc2cc1	96.5	96	45
179	9-ethylcarbazole	86-28-2	C14H13N	c1ccccc2N(CC)c3ccccc32c1	84	92	52
180	*n*-octadecylamine	124-39-1	C18H39N	CCCCCCCCCCCCCCCCCCN	80	93	6
	〈amides〉						
181	acetamide	60-35-5	C2H5NO	CC(=O)N	15	100	0
182	*N*,*N*-dimethylformamide	68-12-2	C3H7NO	O=CN(C)C	24	49	1
183	ε-caprolactam	105-60-2	C6H11NO	N1C(=O)CCCCC1	87.5	99	0
184	acetanilide	103-84-4	C8H9NO	c1cc(NC(=O)C)ccc1	79	100	51
	〈nitriles〉						
185	acetonitrile	75-05-8	C2H3N	CC#N	8	100	0
186	isobutyronitrile	78-82-0	C4H7N	N#CC(C)C	76	96	0
187	2-propenenitrile	107-13-1	C3H3N	C=CC#N	97	96	47
	〈nitro compounds〉						
188	nitromethane	75-52-5	CH3NO2	CN(=O)=O	78	84	2
189	nitroethane	79-24-3	C2H5NO2	CCN(=O)=O	31	63	0.5
190	1-methyl-2-nitrobenzene	88-72-2	C7H7NO2	c1c(N(=O)=O)c(C)ccc1	98	100	0

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
191	1-methyl-3-nitrobenzene	99-08-1	C7H7NO2	c1c(N(=O)=O)cc(C)cc1	67	96	0
192	1-methyl-4-nitrobenzene	99-99-0	C7H7NO2	c1c(N(=O)=O)ccc(C)c1	74	93.5	0
193	1,4-dimethyl-2-nitrobenzene	89-58-7	C8H9NO2	c1c(C)c(N(=O)=O)cc(C)c1	99	99.5	0
194	1-nitronaphthalene	86-57-7	C10H7NO2	c1cc2cccc(N(=O)=O)c2cc1	98	100	4
	〈S-containing compounds〉						
195	ethanethiol	75-08-1	C2H6S	CCS	60	87	31
196	dimethyl sulfide	75-18-3	C2H6S	CSC	32.5	30	14
197	dimethylsulfoxide	67-68-5	C2H6OS	CS(=O)C	11	11	8
198	methyl sulfone	67-71-0	C2H6O2S	CS(=O)(=O)C	0	0	0
199	dimethyl sulfate	77-78-1	C2H6O4S	COS(=O)(=O)OC	100	100	0
200	thiophene	110-02-1	C4H4S	S1C=CC=C1	72	76	33
201	ethyl sulfide	352-93-2	C4H10S	CCSCC	36	41	8
202	thiophenol	108-98-5	C6H6S	c1cc(S)ccc1	98.5	100	16.5
203	benzo[*b*]thiophene	95-15-8	C8H6S	c1cc2SC=Cc2cc1	97.5	100	16
204	diphenylsulfide	139-66-2	C12H10S	c1ccccc1Sc2ccccc2	83	100	2.5
	〈halogenated hydrocarbons〉						
205	dichloromethane	75-09-2	CH2Cl2	ClCCl	2	6.5	0
206	trichloromethane	67-66-3	CHCl3	ClC(Cl)Cl	2	8	0
207	tetrachloromethane	56-23-5	CCl4	ClC(Cl)(Cl)Cl	2	8	0
208	bromochloromethane	74-97-5	CH2BrCl	BrCCl	8.5	41.5	0.2
209	bromodichloromethane	75-27-4	CHBrCl2	BrC(Cl)Cl	6.8	36.8	0.2
210	tribromomethane	75-25-2	CHBr3	BrC(Br)Br	41	52	0.2
211	1,1-dichloroethane	75-34-3	C2H4Cl2	CC(Cl)Cl	4.5	16.5	0
212	1,2-dichloroethane	107-06-2	C2H4Cl2	ClCCCl	6.5	10	0
213	1,2-dibromoethane	106-93-4	C2H4Br2	BrCCBr	33	54	0
214	1,1,1-trichloroethane	71-55-6	C2H3Cl3	ClC(Cl)(Cl)C	28	38	5
215	1,1,2-trichloroethane	79-00-5	C2H3Cl3	ClCC(Cl)Cl	3	9.5	0
216	1,1,2,2-tetrachloroethane	79-34-5	C2H2Cl4	ClC(Cl)C(Cl)Cl	6	10.5	0
217	hexachloroethane	67-72-1	C2Cl6	ClC(Cl)(Cl)C(Cl)(Cl)Cl	2.5	12	1

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
218	trichloroethene	79-01-6	C2HCl3	ClC=C(Cl)Cl	16	18	0
219	tetrachloroethene	127-18-4	C2Cl4	ClC(Cl)=C(Cl)Cl	9	12	0.5
220	allyl chloride	107-05-1	C3H5Cl	C=CCCl	97.5	100	9
221	1,3-dichloro-1-propene	542-75-6	C3H4Cl2	ClCC=CCl	85	84	24
222	1,2dichloropropane	78-87-5	C3H6Cl2	ClCC(Cl)C	12	24.5	0
223	hexachloro-1,3-butadiene	87-68-3	C4Cl6	ClC(Cl)=C(Cl)C(Cl)=C(Cl)Cl	2.5	6	0
224	hexachlorobenzene	118-74-1	C6Cl6	C1(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C1Cl	0	1	0
225	hexafluorobenzene	392-56-3	C6F6	c1(F)c(F)c(F)c(F)c(F)c1F	1.5	18.5	0
226	pentachlorobenzene	608-93-5	C6HCl5	c1c(Cl)c(Cl)c(Cl)c(Cl)c1Cl	0	3.5	0
227	1,2,3,4-tetrachlorobenzene	634-66-2	C6H2Cl4	c1c(Cl)c(Cl)c(Cl)c(Cl)c1	1	9	0
228	1,2,4,5-tetrachlorobenzene	95-94-3	C6H2Cl4	c1(Cl)c(Cl)cc(Cl)c(Cl)c1	2	1.5	0
229	1,2,4-trichlorobenzene	120-82-1	C6H3Cl3	c1c(Cl)c(Cl)cc(Cl)c1	16	51	0
230	1,3,5-trichlorobenzene	108-70-3	C6H3Cl3	c1(Cl)cc(Cl)cc(Cl)c1	9	17	0
231	1,2-dichlorobenzene	95-50-1	C6H4Cl2	c1c(Cl)c(Cl)ccc1	26	70.5	0
232	1,3-dichlorobenzene	541-73-1	C6H4Cl2	c1c(Cl)cc(Cl)cc1	12	34	0
233	1,4-dichlorobenzene	106-46-7	C6H4Cl2	c1c(Cl)ccc(Cl)c1	12.5	34	0
234	chlorobenzene	108-90-7	C6H5Cl	c1ccccc1Cl	51	58	0
235	bromobenzene	108-86-1	C6H5Br	c1ccccc1Br	51	70	0
236	fluorobenzene	462-06-6	C6H5F	c1ccccc1F	51	66	0
237	1,2,3,4,5,6-hexachlorocyclohexane	58-89-9	C6H6Cl6	C1(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C1Cl	1	2	0
238	trichloromethylbenzene	98-07-7	C7H5Cl3	c1ccccc1C(Cl)(Cl)Cl	71	80	6
239	trifluoromethylbenzene	98-08-8	C7H5F3	c1ccccc1C(F)(F)F	4.5	9.5	0
240	1-chloro-2-methylbenzene	95-49-8	C7H7Cl	c1c(Cl)c(C)ccc1	61	89	0
241	chloromethylbenzene	100-44-7	C7H7Cl	c1ccccc1CCl	71	80	6
242	octafluoronaphthalene	313-72-4	C10F8	c1(F)c2c(F)c(F)c(F)c(F)c2c(F)c(F)c(F)1	47	47	0
243	4,4'-dichloro-1,1'-biphenyl	2050-68-2	C12H8Cl2	c1cc(Cl)ccc1c2ccc(Cl)cc2	5	9	0
244	DDE	72-55-9	C14H8Cl4	c1cc(Cl)ccc1C(=C(Cl)Cl))c2ccc(Cl)cc2	5	15	0
245	DDT	50-29-3	C14H9Cl5	c1cc(Cl)ccc1C(C(Cl)(Cl)(Cl))c2ccc(Cl)cc2	2	3	0
	〈multifunctional compounds〉						
246	carbon disulfide	75-15-0	CS2	S=C=S	19	21	2

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
247	cyanamide	420-04-2	CH2N2	NC#N	5	10.5	5
248	urea	57-13-6	CH4N2O	NC(=O)N	32	82	0
249	chloroacetic acid	79-11-8	C2H3ClO2	ClCC(=O)O	76	98	1
250	2,2,2-trichloro-1,1-ethanediol	302-17-0	C2H3Cl3O2	ClC(Cl)(Cl)C(O)O	30	78	2
251	1,2,4-triazole	288-88-0	C2H3N3	N1C=NC=N1	1.8	1.8	0
252	trifluoroacetic acid	76-05-1	C2HF3O2	FC(F)(F)C(=O)O	0	2	0
253	thioacetamide	62-55-5	C2H5NS	CC(=S)N	54	99	51
254	2-chloroethanol	107-07-3	C2H5ClO	ClCCO	84	98	1
255	glycine	56-40-6	C2H5NO2	NCC(=O)O	95	96	3
256	ethylene glycol	107-21-1	C2H6O2	OCCO	97	99.5	9
257	2-aminoethanol	141-43-5	C2H7NO	NCCO	93	100	1
258	ethylenediamine	107-15-3	C2H8N2	NCCN	97	96	0
259	1*H*-imidazole	288-32-4	C3H4N2	N1C=CN=C1	68	71.5	2.5
260	pyruvic acid	127-17-3	C3H4O3	CC(=O)C(=O)O	36.5	98	25
261	malonic acid	141-82-2	C3H4O4	OC(=O)CC(=O)O	99	100	14
262	2,2-dichloropropanoic acid	75-99-0	C3H4Cl2O2	CC(Cl)(Cl)C(=O)O	22.5	94.5	1
263	epichlorhydrin	106-89-8	C3H5ClO	C1CC1OCl	94	99	2
264	chloroacetone	78-95-5	C3H5ClO	ClCC(=O)C	51.5	86.5	1.5
265	methyl chloroacetate	96-34-4	C3H5ClO2	ClCC(=O)OC	79	96.5	1
266	2,3-dibromopropanol	96-13-9	C3H6Br2O	BrCC(Br)CO	78.3	82	3.8
267	glycidol	556-52-5	C3H6O2	OCC1OC1	98	99	19
268	2-hydroxypropanoic acid	598-82-3	C3H6O3	CC(O)C(=O)O	52	98	24
269	2-chloroethyl methyl ether	627-42-9	C3H7ClO	COCCCl	87	100	0
270	carbamic acid ethyl ester	51-79-6	C3H7NO2	NC(=O)OCC	48.5	97	0
271	2-methoxyethanol	109-86-4	C3H8O2	COCCO	89	91	2
272	glycerol	56-81-5	C3H8O3	OCC(O)CO	97	98	22
273	maleic acid	110-16-7	C4H4O4	OC(=O)C=CC(=O)O	98	99	92
274	2,3-butanedione	431-03-8	C4H6O2	CC(=O)C(=O)C	40	97	9.5
275	succinic acid	110-15-6	C4H6O4	OC(=O)CCC(=O)O	79.5	100	0.5
276	*L*-tartaric acid	87-69-4	C4H6O6	OC(=O)C(O)C(O)C(=O)O	100	99	87
277	acetone cyanohydrin	75-86-5	C4H7NO	CC(C)(O)C#N	68	100	0

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
278	*L*-aspartic acid	56-84-8	C4H7NO4	NC(C(=O)O)CC(=O)O	100	100	3.5
279	1,4-dioxane	123-91-1	C4H8O2	C1OCCOC1	100	100	1
280	1-allyl-2-thiourea	109-57-9	C4H8N2S	NC(=S)NCC=C	85	90	48.5
281	morpholine	110-91-8	C4H9NO	C1NCCOC1	85.5	86	0.5
282	4-aminobutanoic acid	56-12-2'	C4H9NO2	NCCCC(=O)O	71.5	100	0
283	di(ethylene glycol)	111-46-6	C4H10O3	OCCOCCO	100	98	4.5
284	diethanolamine	111-42-2	C4H11NO2	OCCNCCO	100	99	18.5
285	1,4-diaminobutane	110-60-1	C4H12N2	NCCCCN	52.5	56	2.5
286	2-furaldehyde	98-01-1	C5H4O2	C1=C(C=O)OC=C1	88.5	100	46
287	furfuryl alcohol	98-00-0	C5H6O2	C1=C(CO)OC=C1	93	99	72.5
288	2,4-pentanedione	123-54-6	C5H8O2	CC(=O)CC(=O)C	39	100	28
289	*L*-glutamic acid	56-86-0	C5H9NO4	OC(=O)C(N)CCC(=O)O	81	100	1
290	pentaerythritol	115-77-5	C5H12O4	OCC(CO)(CO)CO	100	99	19
291	tetrachloro-1,2-benzoquinone	2435-53-2	C6Cl4O2	C1(=O)C(=O)C(Cl)=C(Cl)C(Cl)=C(Cl)1	91.5	80.5	49
292	pentachlorophenol	87-86-5	C6HCl5O	c1(O)c(Cl)c(Cl)c(Cl)c(Cl)c1Cl	49	49	41
293	2,3,4,6-tetrachlorophenol	58-90-2	C6H2Cl4O	c1(Cl)c(O)c(Cl)c(Cl)c(Cl)c1	95	86.5	70
294	1-chloro-2,4-dinitrobenzene	97-00-7	C6H3ClN2O4	c1c(N(=O)=O)ccc(Cl)c1N(=O)=O	9	12	0
295	1,2-dichloro-4-nitrobenzene	99-54-7	C6H3Cl2NO2	c1c(N(=O)=O)cc(Cl)c(Cl)c1	3	11	0.5
296	2,3,6-trichlorophenol	933-75-5	C6H3Cl3O	c1c(Cl)c(O)c(Cl)c(Cl)c1	89	89	70
297	2,4,5-trichlorophenol	95-95-4	C6H3Cl3O	c1(O)c(Cl)cc(Cl)c(Cl)c1	90	93	54.5
298	2,4,6-trichlorophenol	88-06-2	C6H3Cl3O	c1(Cl)c(O)c(Cl)cc(Cl)c1	97	93	80
299	2,4,6-trinitrophenol	88-89-1	C6H3N3O7	c1(N(=O)=O)c(O)c(N(=O)=O)cc(N(=O)=O)c1	97	96	41
300	1,4-benzoquinone	106-51-4	C6H4O2	C1(=O)C=CC(=O)C=C1	98	96	98
301	2,3-dichlorophenol	576-24-9	C6H4Cl2O	c1cc(O)c(Cl)c(Cl)c1	100	99	86
302	2,4-dichlorophenol	120-83-2	C6H4Cl2O	c1c(O)c(Cl)cc(Cl)c1	99	95	87.5
303	2,6-dichlorophenol	87-65-0	C6H4Cl2O	c1c(Cl)c(O)c(Cl)cc1	98	95	73
304	1-chloro-4-nitrobenzene	100-00-5	C6H4ClNO2	c1c(Cl)ccc(N(=O)=O)c1	3	16	0
305	1-chloro-2-nitrobenzene	88-73-3	C6H4ClNO2	c1c(Cl)c(N(=O)=O)ccc1	24	60.5	0
306	2-chloro-4-nitrophenol	619-08-9	C6H4ClNO3	c1c(N(=O)=O)cc(Cl)c(O)c1	100	100	100
307	4-chloro-2-nitrophenol	89-64-5	C6H4ClNO3	c1c(Cl)cc(N(=O)=O)c(O)c1	81	80	77
308	1,3-dinitrobenzene	99-65-0	C6H4N2O4	c1c(N(=O)=O)cc(N(=O)O)cc1	3	5	0

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
309	pentachloronitrobenzene	82-68-8	C6Cl5NO2	c1(Cl)c(Cl)c(Cl)c(N(=O)=O)c(Cl)c1Cl	3.5	2.5	0
310	4-chlorophenol	106-48-9	C6H5ClO	c1c(Cl)ccc(O)c1	97	98.5	82
311	2-fluorophenol	367-12-4	C6H5FO	c1cc(F)c(O)cc1	100	100	84
312	4-fluorophenol	371-41-5	C6H5FO	c1c(F)ccc(O)c1	97	97	82.5
313	2-nitrophenol	88-75-5	C6H5NO3	c1c(N(=O)=O)c(O)ccc1	86	95	89
314	4-nitrophenol	100-02-7	C6H5NO3	c1c(N(=O)=O)ccc(O)c1	100	100	60
315	2-amino-4-chlorophenol	95-85-2	C6H6ClNO	Oc1ccc(Cl)cc1N	100	100	87.5
316	1,2-benzenediol	120-80-9	C6H6O2	c1c(O)c(O)ccc1	97	99	84
317	hydroquinone	123-31-9	C6H6O2	c1c(O)cc(O)cc1	94	96	93
318	1,2,3-benzenetriol	87-66-1	C6H6O3	c1(O)c(O)c(O)ccc1	94	96	93
319	1,3,5-benzenetriol	108-73-6	C6H6O3	c1(O)cc(O)cc(O)c1	80	95	69
320	4-aminophenol	123-30-8	C6H7NO	c1c(N)ccc(O)c1	94.5	96	58
321	citric acid	77-92-9	C6H8O7	OC(=O)CC(O)(C(=O)O)CC(=O)O	99	100	76
322	nitrilotriacetic acid	139-13-9	C6H9NO6	OC(=O)CN(CC(=O)O)CC(=O)O	96	98	86
323	ethyl acetoacetate	141-97-9	C6H10O3	CC(=O)CC(=O)OCC	50	97	23
324	hexanedioic acid	124-04-9	C6H10O4	OC(=O)CCCCC(=O)O	77.5	99	0
325	diethyl oxalate	95-92-1	C6H10O4	CCOC(=O)C(=O)OCC	44	100	31.5
326	4-hydroxy-4-methyl-2-pentanone	123-42-2	C6H12O2	CC(=O)CC(C)(O)C	64	100	10
327	*L*-cystine	56-89-3	C6H12N2O4S2	OC(=O)C(N)CSSCC(N)C(=O)O	73	100	38
328	*L*-leucine	61-90-5	C6H13NO2	CC(C)CC(N)C(=O)O	80	99	2
329	2-butoxyethanol	111-76-2	C6H14O2	CCCCOCCO	85	99	3
330	tri(ethylene glycol)	112-27-6	C6H14O4	OCCOCCOCCO	100	100	4
331	*D*-mannitol	69-65-8	C6H14O6	OCC(O)C(O)C(O)C(O)CO	98	96.5	72
332	triethanolamine	102-71-6	C6H15NO3	OCCN(CCO)CCO	98	95	11
333	hexamethylenediamine	124-09-4	C6H16N2	NCCCCCCN	65	100	0.5
334	4-chlorophenyl isocyanate	104-12-1	C7H4ClNO	c1c(Cl)ccc(N=C=O)c1	95	94	41
335	phenyl isocyanate	103-71-9	C7H5NO	c1ccccc1N=C=O	95	97	64
336	phenyl isothiocyanate	103-72-0	C7H5NS	c1ccccc1N=C=S	94.5	96	5
337	benzothiazole	95-16-9	C7H5NS	S1c2ccccc2N=C1	92	94	72
338	4-nitrobenzoic acid	62-23-7	C7H5NO4	c1c(N(=O)=O)ccc(C(=O)O)c1	97.5	92	0
339	2-chlorobenzoic acid	118-91-2	C7H5ClO2	c1c(Cl)c(C(=O)O)ccc1	71	71	0.5

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
340	2-hydroxybenzaldehyde	90-02-8	C7H6O2	c1c(O)c(C=O)ccc1	97	99	79.5
341	2-hydroxybenzoic acid	69-72-7	C7H6O3	c1c(O)c(C(=O)O)ccc1	99	99	100
342	2,4-dihydroxybenzoic acid	89-86-1	C7H6O4	c1c(O)cc(O)c(C(=O)O)c1	100	100	84
343	3,4-dihydroxybenzoic acid	99-50-3	C7H6O4	c1c(O)c(O)cc(C(=O)O)c1	97.5	97.5	90.5
344	3-aminobenzotrifluoride	96-16-8	C7H6F3N	c1c(C(F)(F)F)cc(N)cc1	86.8	91.8	73.5
345	2,4-dinitrotoluene	606-20-2	C7H6N2O4	c1c(C+)c(N(=O)=O)cc(N(=O)=O)c1	99	100	0
346	2-methyl-4,6-dinitrophenol	534-52-1	C7H6N2O5	c1(N(=O)=O)cc(N(=O)=O)c(O)c(C)c1	85	96.5	41
347	4-chloro-1-methyl-2-nitrobenzene	89-59-8	C7H6ClNO2	O=N(=O)c1cc(Cl)ccc1C	88	92	0
348	4-chloro-2-methylphenol	1570-64-5	C7H7ClO	c1c(Cl)cc(C)c(O)c1	83	100	65
349	4-chloro-3-methylphenol	59-50-7	C7H7ClO	c1c(Cl)c(C)cc(O)c1	83	100	69
350	2-aminobenzoic acid	118-92-3	C7H7NO2	c1c(N)c(C(=O)O)ccc1	99	100	90.5
351	2-hydroxybenzamide	65-45-2	C7H7NO2	c1c(O)c(C(=O)N)ccc1	100	98	82
352	1-methoxy-2-nitrobenzene	91-23-6	C7H7NO3	c1c(OC)c(N(=O)=O)ccc1	100	100	0.5
353	4-methyl-2-nitrophenol	119-33-5	C7H7NO3	c1c(C)cc(N(=O)=O)c(O)c1	84	97	50
354	2,4-dihydroxybenzamide	3147-45-3	C7H7NO3	c1cc(O)cc(O)c1C(=O)N	99	100	100
355	2-methoxyphenol	90-05-1	C7H8O2	Oc1ccccc1OC	96	97	67
356	4-methoxyphenol	150-76-5	C7H8O2	Oc1ccc(OC)cc1	99	98.5	83
357	2-hydroxybenzyl alcohol	90-01-7	C7H8O2	c1cccc(O)c1CO	100	100	76
358	2-methoxyaniline	90-04-0	C7H9NO	Nc1ccccc1OC	100	100	79
359	phthalic acid	88-99-3	C8H6O4	OC(=O)c1ccccc1C(=O)O	100	100	1
360	2,4-dichlorophenoxyacetic acid	94-75-7	C8H6Cl2O3	c1cc(Cl)cc(Cl)c1OCC(=O)O	95.5	100	66
361	2,4,5-trichlorophenoxyacetic acid	93-76-5	C8H5Cl3O3	c1c(Cl)c(Cl)cc(Cl)c1OCC(=O)O	98	100	10.5
362	2-hydroxy-4-methylbenzoic acid	50-85-1	C8H8O3	c1c(C)cc(O)c(C(=O)O)c1	82	94	65.5
363	α-hydroxybenzeneacetic acid	90-64-2	C8H8O3	c1ccc(C(O)C(=O)O)cc1	100	100	14
364	2-ethyl-1,3-hexanediol	94-96-2	C8H18O2	CCCC(O)C(CO)CC	76.5	100	4
365	di(ethylene glycol)butyl ether	112-34-5	C8H18O3	CCCCOCCOCCO	88	99	2
366	8-hydroxyquinoline	148-24-3	C9H7NO	c1c(O)c2ncccc2cc1	85.5	90.5	64.5
367	*N*-benzoylglycine	495-69-2	C9H9NO3	c1ccccc1C(=O)NCC(=O)O	98	100	1
368	4-chloro-2-methylphenoxyacetic acid	94-74-6	C9H9ClO3	c1cc(Cl)cc(C)c1OCC(=O)O	85	97	62
369	4-hydroxy-3,5-dimethoxybenzoic acid	530-57-4	C9H10O5	c1c(OC)c(O)c(OC)cc1C(=O)O	100	100	71.5
370	*N'*-(3,4-dichlorophenyl)-*N*-methoxy-	330-55-2	C9H10Cl2N2O2	c1c(Cl)c(Cl)ccc1NC(=O)N(C)OC	94	91	65

Table 2 (continued).

No.	Compound	CAS no.	Formula	SMILES	COD%		
					$K_2Cr_2O_7$	$K_2Cr_2O_7+Ag^+$	$KMnO_4$
	N-methyl-urea						
371	4-dimethylaminobenzaldehyde	100-10-7	C9H11NO	c1cc(N(C)C)ccc1C=O	85	87	65
372	*N'*-(4-chlorophenyl)-*N*-methoxy-*N*-methyl-urea	1746-81-2	C9H11ClN2O2	c1cc(Cl)ccc1NC(=O)N(C)OC	93.5	95	53
373	dimethyl phthalate	131-11-3	C10H10O4	c1(C(=O)OC)ccccc1C(=O)OC	100	100	0
374	dimethyl terephthalate	120-61-6	C10H10O4	c1cc(C(=O)OC)ccc1C(=O)OC	100	100	0
375	2-methoxy-4-allylphenol	97-53-0	C10H12O2	c1c(OC)c(O)ccc1CC=C	91	93	78
376	2-methoxy-4-propenylphenol	97-54-1	C10H12O2	c1c(OC)c(O)ccc1C=CC	86	96	69
377	sebacic acid	111-20-6	C10H18O4	OC(=O)CCCCCCCCC(=O)O	90	100	0.5
378	*L*-tryptophan	73-22-3	C11H12N2O2	c1cc2NC=C(CC(N)C(=O)O)c2cc1	94	94	74.5
379	10*H*-phenothiazine	92-84-2	C12H9NS	c1cc2Nc3ccccc3Sc2cc1	94.5	96	50
380	*N'*-(4-chlorophenyl)-*N*-methyl-*N*-(1-methyl-2-propynyl)-urea	3766-60-7	C12H13ClN2O	c1cc(Cl)cc1NC(=O)N(C)C(C)C#C	77	87	13
381	diethyl phthalate	84-66-2	C12H14O4	CCOC(=O)c1ccccc1C(=O)OCC	71	100	0
382	*N,N'*-diphenylurea	102-07-8	C13H12N2O	c1ccccc1NC(=O)Nc2ccccc2	93	94	60
383	1,5-diphenylcarbazide	140-22-7	C13H14N4O	c1ccccc1NNC(=O)NNc2ccccc2	94.5	92	35
384	anthraquinone	84-65-1	C14H8O2	c1cc2C(=O)c3ccccc3C(=O)c2cc1	32	37	0
385	benzil	134-81-6	C14H10O2	c1ccccc1C(=O)C(=O)c2ccccc2	97	98.5	0
386	benzoin	579-44-2	C14H12O2	c1ccccc1C(O)C(=O)c2ccccc2	97.5	100	3
387	diallyl phthalate	131-17-9	C14H14O4	C=CCOC(=O)c1ccccc1C(=O)OCC=C	100	100	27
388	di-*n*-butylphthalate	84-74-2	C16H22O4	CCCCOC(=O)c1ccccc1C(=O)OCCCC	79.5	84	0
389	sebacic acid di-*n*-butyl ester	109-43-3	C18H34O4	CCCCOC(=O)CCCCCCCCC(=O)OCCCC	75	91.5	0
390	dicyclohexyl phthalate	84-61-7	C20H26O4	C1CCCCC1OC(=O)c2ccccc2C(=O)OC3CCCCC3	91	97	0

Descriptors

To characterize the compounds, a set of 30 molecular descriptors listed in Table 3 was utilized. The descriptors were divided into two types; physico-chemical descriptors, D1, and structural descriptors, D2, which were analyzed separately. They were selected based on general rules of thumb concerning the effects of chemical structure on chemical oxidizability and/or biodegradability (see Table1). Electronic parameters, E_{HOMO}, E_{LUMO}, μ , q_m^+ and q_m^-, were calculated with the semiempirical quantum chemical AM1 model using the MOPAC program (MOPAC 93, CambridgeSoft). Oxidation reactions are intimately related to the gain

Table 3. Descriptors used to model chemical oxidizabilities for organic compounds.

Set D1, Physico-chemical descriptors		
1	MW	Molecular weight
2	MV	Liquid state molar volume at 298K
3	$^0\chi$	Zero-th order molecular connectivity index
4	$^1\chi$	First order molecular connectivity index
5	E_{HOMO}	Highest occupied molecular orbital energy
6	E_{LUMO}	Lowest occupied molecular orbital energy
7	μ	Dipole moment
8	q_m^+	Absolute charge of the most positive atom
9	q_m^-	Absolute charge of the most negative atom
10	log P	Partition coefficient in 1-octanol/water
Set D2, Structural descriptors		
11	N_C	No. of C atoms
12	N_H	No. of H atoms
13	N_O	No. of O atoms
14	N_N	No. of N atoms
15	N_S	No. of S atoms
16	N_X	No. of halogen atoms
17	$N_{C=C,C\equiv C}$	No. of double or triple bonds between C atoms
18	$N_{C=N}$	No. of C=N bonds
19	$N_{C\equiv N}$	No. of C $\equiv$ N bonds
20	N_{ar}	No. of aromatic bonds
21	N_R	No. of rings
22	$N_{C=O/C=S}$	No. of -C=O or -C=S groups
23	N_{COO}	No. of -COO- groups
24	N_{CHO}	No. of -CHO groups
25	N_{COOH}	No. of -COOH groups
26	$N_{NH2/NH}$	No. of -NH_2 or -NH- groups
27	$N_{OH/SH}$	No. of -OH or -SH groups
28	$N_{-O-/-S-}$	No. of -O- or -S- groups
29	N_{NO2}	No. of -NO_2 groups
30	$N_{SO/SO2}$	No. of SO or SO_2 groups

and loss of electrons by the interacting species. The E_{HOMO} and/or E_{LUMO} can be expected that such parameters are an indication of the energy required for a compound to donate an electron to an oxidant (Rorije & Peijnenburg 1996) and most common in toxicological QSAR. According to Parr and Pearson (1983), two parameters which are calculated from the HOMO and LUMO energies, the electronegativity EN and the hardness H, may be useful entities for characterizing molecular oxidizability:

$$EN = -(E_{HOMO} + E_{LUMO}) / 2 \quad (2)$$

$$H = -(E_{HOMO} - E_{LUMO}) / 2 \quad (3)$$

These parameters can be included implicitly in this study by applying the PLS calculations from the original E_{HOMO} and E_{LUMO}. The observed log P values were taken from a databank package, LOGKOW (2nd ed., 1993).

Partial least squares (PLS) calculations

The method of PLS is a generalized regression method that allows to build a predictive model between two matrices of data (Wold et al 1984). In principle, PLS is based on two PCA models that are rotated to maximize the overlap between the target property (Y) and the predictor (X) matrices. The advantage of the PLS treatment of data is that the obtained model involves orthogonal scales that eliminates the problems arising from the collinearity of descriptors.

Artificial neural network (ANN) calculations

The ANN with one hidden layer, a varying number of hidden-layer neurons, and three output neurons representing three different COD observations were optimized by using an in-house developed program (Suzuki & Ishida 1995). The architecture of the model used is shown in Fig. 1. The inputs consist of the above two sets of descriptors. All inputting descriptor x were transformed to normalized values x' between 0.05 and 0.95 using the following transformation:

$$x_i' = 0.9 \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} + 0.05 \quad (4)$$

where x_i is the value of the *ith* input value, x_{min} and x_{max} are its minimum and maximum data over the data set. The same range-scaling formula was applied to the training values of experimental COD data. The ANNs are fully connected, i.e., no short-cut connections between layers or pruning were used. Bias units were attached to hidden and output layers. The ANNs were trained using back-propagation algorithm.

The statistical quality of the modeling results for both training and prediction sets was evaluated based on the following two parameters: squared correlation coefficient (r^2) and the root-mean-square error (*RMSE*):

$$r^2 = 1 - \sum_{i=1}^{n} (y_i - y_i^{fit})^2 / \sum_{i=1}^{1} (y_i - y_{mean})^2 \quad (5)$$

$$RMSE = \sqrt{\sum_{i-1}^{n} (y_i - y_i^{fit})^2 / n} \quad (6)$$

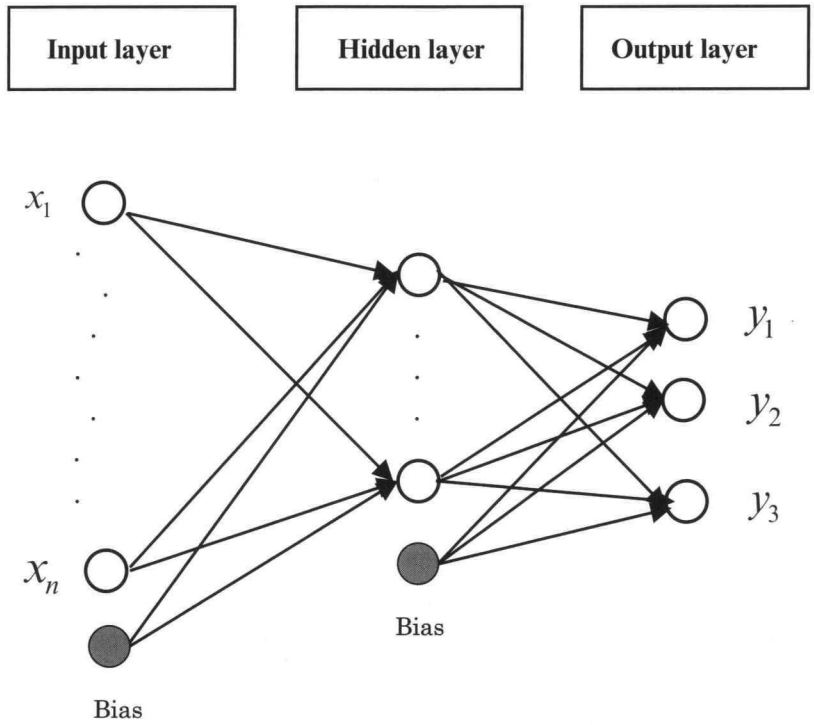

Fig 1. Architecture of neural network model for chemical oxidizability modeling.

where y_i is the experimental target value (COD) for the ith compound, y_{mean} denotes the mean value of y_i, and y_i^{fit} is the calculated target value.

3. Results and Discussions

The scheme for developing the prediction models for COD values is shown in Fig. 2. The PLS modeling using 10 kinds of physico-chemical descriptors for 292 compounds from the entire data set of 390 compounds did not gave any statistically significant models for the three different COD data, respectively (r values in the range of 0.597-0.618). However, this scheme gave some statistically significant regression models for homologous series of the compound as follows;

Hydrocarbons:

$$y_1 = -0.00159z_1 + 0.0474z_2 + 0.161\,z_3 + 0.56 \quad (7)$$

n=37, r^2=0.853, $RMSE$ = 0.12

Halogenated hydrocarbons:

$$y_2 = 0.167z_1 + 0.0984z_2 + 0.155z_3 + 0.114z_4 + 0.0874z_5 + 0.078z_6 + 0.43 \quad (8)$$

n=33, r^2=0.853, $RMSE$ = 0.10

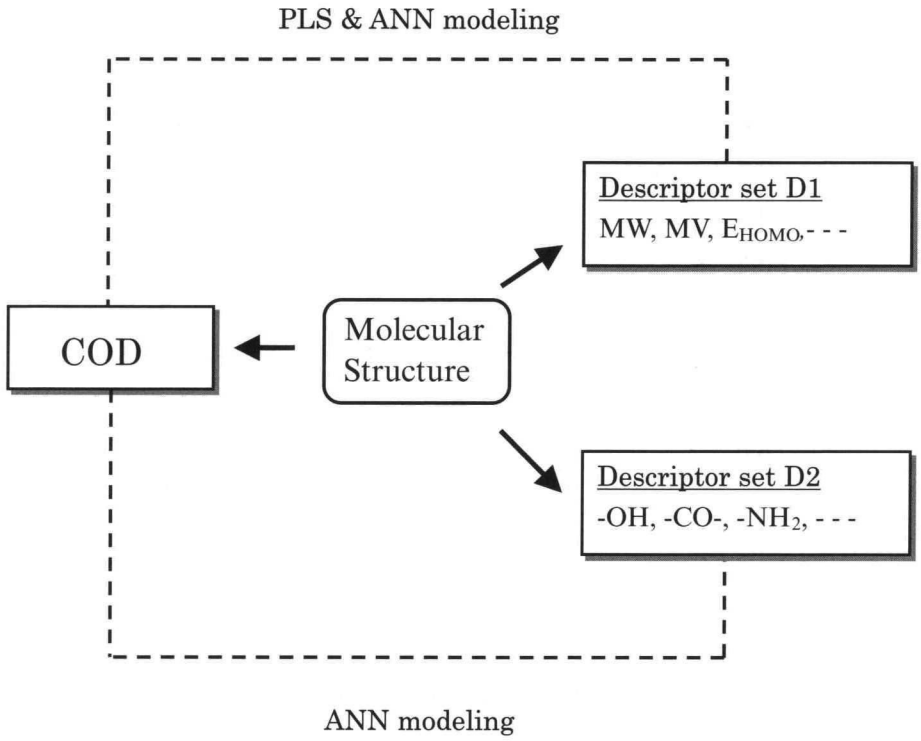

Fig 2. Prediction scheme for COD values of a variety of organic compounds.

where y_i are target COD values and zs are latent variables. Each latent variable is a linear combination of the original 10 physico-chemical descriptors.

For the ANN modeling, initial attempts were made to select a better set of descriptors to this COD problem, with a subset of 292 compounds out of 390 compounds. The subset was split into two parts: 146 compounds for the training and 146 compounds for prediction sets under the guidance that the structural variety of both sets is similar with regard to the relative portions of the major compound classes. Plots of the observed $COD_{K2Cr2O7}$ vs $COD_{K2Cr2O7\text{-}Ag+}$ and $COD_{K2Cr2O7}$ vs COD_{KMnO4} for the 146 training set compounds showed a considerable linear correlation exists between $COD_{K2Cr2O7}$ and $COD_{K2Cr2O7\text{-}Ag+}$, while there is only a low correlation between $COD_{K2Cr2O7}$ and COD_{KMnO4}.

The model performance for the training and test sets with descriptor sets D1 and D2 are shown in Figs. 3 and 4, respectively, as a function of training epochs. From the results, it can be confirmed that the ANN model gave a global model for the three different COD values with a reasonable accuracy for a wide range of chemical structures.

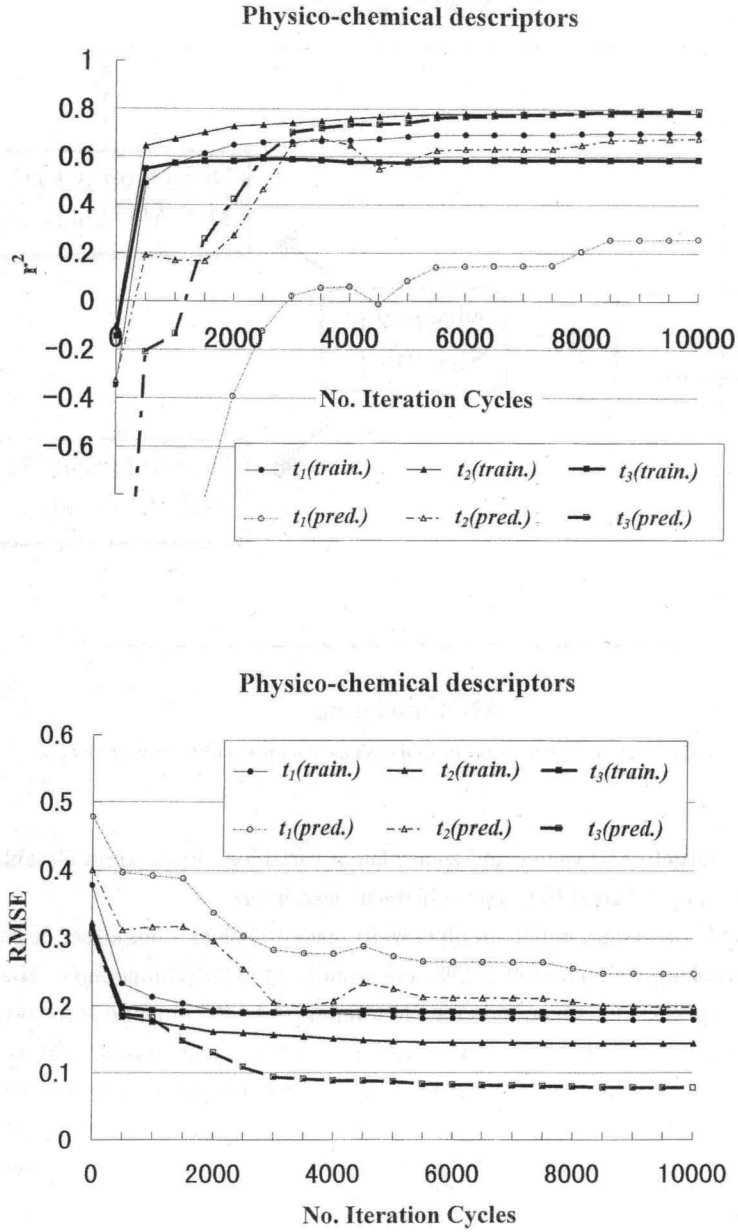

Fig 3. Neural network performance for the model with physico-chemical descriptors as a function of the training epochs.

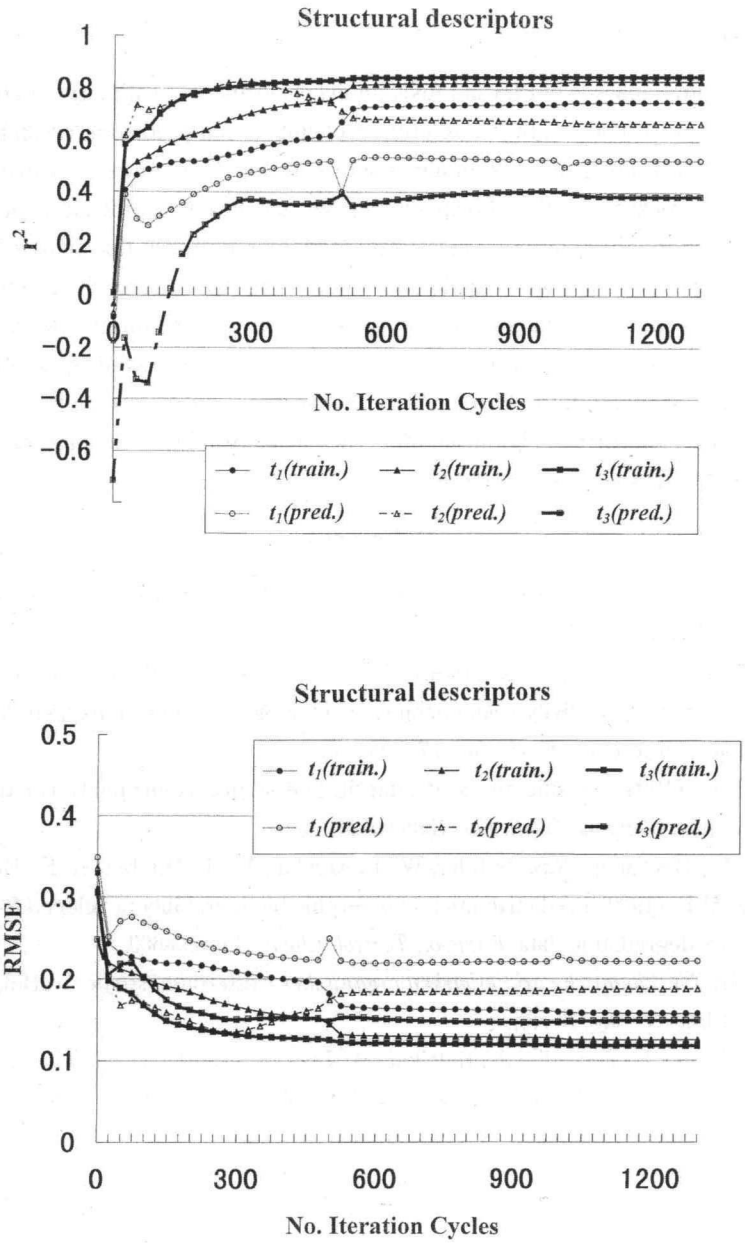

Fig 4. Neural network perfomance for the model with stuctural descriptors as a function of the training epochs

4. Conclusions

The comparative analysis of PLS and ANN model performances with different training and prediction sets demonstrates, that a different recognition and prediction capabilities of two modeling schemes. For some homologous series of the compounds (e.g. hydrocarbons, halogenated hydrocarbons), the PLS approach gave statistically significant models for COD (values based on the three different measuring methods), however, this approach could not establish a single model for all studied compounds. The ANN approach gave much better results than PLS approach and could obtain global models for the three different COD data used. The predictive ability of the proposed methods were tested for the testing set compounds not included in the training data set and fairly good agreement with observed COD values was confirmed. Significant difference in adaptability of the proposed models to three different COD values was observed.

References

Babeu, L., Vaishnav, D.D. (1987) Prediction of biodegrability for selected organic chemicals. *J. Ind. Microbiol.* **2**, 107-115.

Dearden, J. C. , Nicholson, R. M.(1986) The prediction of biodegrability by the use of quantitative structure-activity relationships: correlation of biological oxygen demand with atomic charge difference. *Pestic. Sci.* **17**, 305-310.

Dearden, J. C. (1994) Applications of quantitative structure-property relationships to pharmaceutics. *Chemom. Intell. Lab. Systems* **24**, 77-87.

Howard, P. H., Boethling, R.S., Stiteler, W.M., Meylan, W.M., Hueber, A. E., Beauman, J., Larosche, M.E. (1992) Predictive model for aerobic biodegrability developed from a file of evaluated biodegradation data. *Environ. Toxicol. Chem.,* **11**, 593-603.

Janicke, W.(1983) *Chemische oxidierbarkeit organischer wasserinhaltsstoffe.* WaBoLu-Berichte, Dietrich Reimer Verlag, Berlin.

Lyman,W., Nelson, L., Partridge, L., Kalelkar , A., Everett, J., Allen, D., Goodier, L.L., Pollack, G.(1974) "Survey Study to Select a Limited Number of Hazardous Materials to Define Amelioration Requirements," U.S. Coast Guard, CG-D-46-75.

Niemi, G. J., Veith, G. D., Regal, R.R., Vaishnav, D.D. (1987) Structural features associated with degradable and persistent chemicals. *Environ. Toxicol. Chem.* **6**, 515- 527.

Parr, R. G., Pearson, R. G.(1983) Absolute hardness: companion parameter to absolute electronegativity. *J. Am. Chem. Soc.* **105**, 7512-7516.

Pitter, P.(1976) Determination of biological degradability of organic substances," *Water Res.,* **10**, 231-235.

Rorije, E., Peijnenburg, W.J.G.M.(1996) QSARs for oxidation of phenols in the aqueous environment, suitable for risk assessment. *J. Chemometrics* **10**, 79-93.

Scow, K.M. (1982) *Rate of biodegradation.* In:Lyman, W.J., Reehl, W.F., Rosenblatt, D. H.(eds)

Handbook of chemical property estimation methods, McGraw-Hill, New York, Chap. 9.

Suzuki, T., Ishida, M. (1995) Neural network techniques applied to predict flammability limits of organic compounds. *Fire & Materials* **19**, 179-189.

Suzuki, T., Toda, T., Ishida, M.(1997) Prediction of chemical oxidizabilities of organic compounds by QSAR techniques. *J. Pharm. Pharmacol.* **49**(Suppl. 4), 107.

Tabak, H.H., Govind, R. (1993) Prediction of biodegradation kinetics using a nonlinear group contribution method. *Environ. Toxicol. Chem.* **12**, 251-260.

Weininger, D., Weininger, J. L.(1990) In *Quantitative drug design*, (ed. C. A. Ramsden), Vol. 4 of *Comprehensive medicinal chemistry. The rational design, mechanistic study and therapeutic application of chemical compounds* (ed. C. Hansch, P.G. Sammers, J.B. Taylor), pp.59-82. Pergamon Press, Oxford.

Wold, S., Albano, C., Dunn III, W.J., Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W., Sjöström, M.(1984) *Multivariate data analysis in chemistry.* In:Kowalski, B.R. (ed) *Chemometrics, mathematics and statistics in chemistry*, Reidel, Dordrecht, pp. 17-95.

Zakarya, D., Belkhadir, M., Fkih-Tetouani, S. (1993) Quantitative structure-biodegrability relationships (QSBRs) using modified autocorrelation method (MAM). *SAR QSAR Environ. Res.* **1**, 21-27.

Zitko, V. (1991) Prediction of biodegrability of organic chemicals by an artificial network. *Chemosphere* **23**, 305-312.

Zupan, J., Gasteiger, J. (1993) *Neural networks for chemists: an introduction*. VCH,Weinheim.

要　旨

鈴木孝弘：有機化合物の酸化分解特性の QSAR の手法による予測

化学物質の環境中での動態を評価するパラメータの一つとして COD（Chemical Oxygen Demand; 化学的酸素要求量）がある．本研究では，分子設計の段階で物質の化学構造の情報のみから，その COD 値を精度よく予測する手法の開発を目的とした．390 種の化合物について報告されている測定方法の異なる 3 種類の COD 値を目的変数とした．記述子には，化学物質の構造から得られ，酸化反応の機構と何らかの関係があると推定される化合物側の基礎的な物性である分子の体積，極性，電子密度などに関係するパラメータ 10 種と分子中の特定の原子や結合，原子団などの有無と数を示すパラメータ 20 種の 2 セットを用いた．モデル化には，QSAR（薬物の定量的構造活性相関）の手法のうちで，記述子から潜在的な因子を抽出して目的変数との線形の相関を検討する PLS（Partial Least Squares）回帰分析と非線形の相関をみるためにニューラルネットワーク（ANN）を適用した．PLS によるモデル化では，個別の炭化水素，アルコール，エステルなど同族列の化合物群に対しては，相関のよいモデルを構築できたが，全体の化合物を単一のモデ

ルで表わすことが出来なかった．一方，ANN によるモデル化では，分子の大きさや構造のパラメータを用いて，多種多様な全体の化合物を一つのモデルで精度よく表現できることがわかった．