Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Medicinal Chemistry 38 (2003) 939–947
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
www.elsevier.com/locate/ejmech

Original article

Quantitative relationship between rat intestinal absorption and Abraham descriptors

Yuan H. Zhao [a], Michael H. Abraham [a,*], Anne Hersey [b], Chris N. Luscombe [b]

[a] *Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK*
[b] *Computational and Structural Sciences (CASS), GlaxoSmithKline, Gunnels Wood Road, Stevenage, Herts SG1 2NY, UK*

Received 24 June 2003; accepted 12 July 2003

Abstract

Literature data on the intestinal absorption of 158 drug and drug-like compounds in rats have been collected, and Abraham descriptors for the set of drugs have been calculated using the method of Platts and Abraham et al. Results show that there is a significant relationship between rat intestinal absorption and the Abraham descriptors. In agreement with the human intestinal absorption model, the dominant descriptors in the rat model are the drug hydrogen bond acidity and basicity. In order to compare the absorption models in humans and rats, the absorption model developed from rats was used to predict the absorption in humans. The rat intestinal absorption model is similar to the human absorption model, and data on rats can effectively be used to predict human intestinal absorption.

Keywords: Rat intestinal absorption; Kinetic rate; Abraham descriptor; Correlation

1. Introduction

Combinatorial chemistry and high throughput screening are widely used in drug research and discovery, e.g. for prediction of oral absorption, distribution, metabolism and elimination (ADME). Several structure–activity relationship (SAR) and quantitative structure–activity relationship (QSAR) models have been developed to establish the relationship between absorption and molecular descriptors, such as lipophilicity [1], absorption potential (AP) [2,3], molecular polar surface area (PSA) [4–6] and various linear free energy relationship (LFER) descriptors [7–10].

Among these models, hydrogen bond donor and acceptor descriptors have been shown to be important in the modelling of human absorption by SAR or QSAR analysis. Lipophilicity, commonly expressed by the octanol/water partition coefficient as log P_{oct}, has been successfully used to predict passive drug absorption in vivo for series of homologous compounds [1]. Log P_{oct} itself is known to encode mainly hydrogen bond basicity and solute size [11,12]. To correct for low solubility, Dressman et al. [2] and Macheras et al. [3] introduced the concept of absorption potential (AP). With this approach, log P_{oct} is corrected for the molar fraction of nonionized species at pH 6.5 (F_{non}), the solubility of the nonionized species in water (S_w), the volume of the luminal contents (V_L) and the dose administered (X_O).

$$AP = \log\left(P_{oct} \times F_{non} \times \frac{S_w \times V_L}{X_O}\right) \quad (1)$$

Both Dressman et al. and Macheras et al. found a sigmoidal relationship between the fraction absorbed in humans and the AP for seven chemically different drug compounds.

The Lipinski 'rule of 5' uses four descriptors to produce an alert for compounds likely to be poorly absorbed [13]. The four descriptors are: log P_{oct}, molecular weight, and the number of H-bond donor and acceptor atoms in the molecule; of these four, two are directly related to hydrogen bond properties. The alert has a primary value in identifying problem compounds, although compounds passing the alert can still prove troublesome in experimental studies. Polar surface area

* Corresponding author.
E-mail address: m.h.abraham@ucl.ac.uk (M.H. Abraham).

0223-5234/03/$ - see front matter © 2003 Éditions scientifiques et médicales Elsevier SAS. All rights reserved.
doi:10.1016/j.ejmech.2003.07.005

(PSA), defined as the area occupied by nitrogen and oxygen atoms, and hydrogen atoms attached to these heteroatoms, has been used by Palm and Clark and coworkers [4–6] to correlate human absorption. However, the correlation is good only for 20 model compounds [4] and the fit was poor when the model was applied to a larger data set of compounds [6].

The LFER developed by the Abraham group contains five main descriptors: solute excess molar refractivity, dipolarity/polarisability, summation hydrogen bond acidity and basicity, and the McGowan characteristic volume [11,12]. We have applied the LFER equation, using calculated descriptors, to human oral absorption of 241 drug and drug-like compounds with a wide range of physicochemical properties [7–10]. The result showed that the dominant descriptors were those for hydrogen bond basicity and hydrogen bond acidity, which was consistent with previously reported results [4–6,13,14]. Regression analysis was also performed between absorption and log P_{oct} or PSA. In comparing these models, LFER or log P_{oct} methods seem better than PSA for the correlation of human oral absorption [9].

Although there has been considerable work done on establishing quantitative relationships between human absorption and molecular descriptors, only the study of Chiou and Barve has compared human and rat absorption for a large data set (80 compounds) [15–18]. Rat is one of the animals commonly used for oral absorption in preclinical studies. These studies are carried out on the assumption that the results obtained may be successfully extrapolated to humans. Chiou and Barve [15] reported an excellent overall correlation ($r^2 = 0.97$) between human and rat absorption for 64 compounds with wide physicochemical and pharmacological properties. In contract, the overall correlation was relatively poor ($r^2 = 0.512$) between human and dog absorption for 43 compounds [16]. They concluded that rat may generally serve as a reliable animal model to predict or study drug absorption in humans.

Recently, we have obtained and evaluated 111 rat absorption data for drug and drug-like compounds from the literature [19]. Combining this set with the Chiou and Barve data set results in rat absorption data for 158 drug compounds. From this set 98 drug compounds for which both human and rat absorption data were available were selected for correlation analysis between the human and rat absorption. The results showed that the extent of absorption in these two species was similar, with a standard deviation of 11% [19]. The aim of this study is, firstly to develop a QSAR model for rat intestinal absorption using the LFER equation of Abraham, and secondly to investigate the similarity in absorption models between humans and rats. Values of log P_{oct} and the polar surface area were also calculated in this paper for use in QSAR analyses.

2. Chemistry

2.1. Rat intestinal absorption data

The rat intestinal absorption dosed orally by gavage was collected and evaluated from 111 sources of literature [19]. Absorption data on other compounds were taken from Chiou and Barve [15,16]. Methods of obtaining qualified oral absorption have previously been reported [8,15]. One of the best ways of obtaining oral absorption data is from the ratio of cumulative urinary excretion of drug and drug-related materials following oral and intravenous administration, this method being especially used for low absorption. Table 1 lists the absorption data evaluated mostly by this method and also some absorption data obtained from the percentage of urinary and biliary excretion.

2.2. Human intestinal absorption data

Data on human intestinal absorption dosed orally was collected and evaluated from 244 sources of literature; details have been published previously [8].

2.3. Physicochemical descriptors

Abraham molecular descriptors were calculated using a Unix program derived from the work of Platts and Abraham et al. [20]. The program is written to read molecular structures as SMILES strings. After calculation of the solvation descriptors, a code is given by the program for each compound as an indication of the quality of the parameter calculations. Calculated values of log P_{oct} were obtained by use of the ClogP program for Windows software (Biobyte version 4.0, Claremont, CA). Polar molecular surface area (PSA) was calculated by use of the SAVOL program (Tripos Inc. 1699 South Hanley Road St. Louis, MO). The 3D structure of a compound was generated from SMILES using CONCORD. After energy minimisation, the polar surface area for each compound was calculated.

2.4. Statistical analysis

The data set was analysed using EXCEL 97. Stepwise regression analysis was used to determine the most significant descriptors. Partial least-squares regression (PLS) was also carried out on the data set using Simca-P (version 9.0). The regression coefficients were obtained by least-squares regression analysis. For each regression, the following information is provided: number of observations used in the analysis (n), square of the coefficient of determination (r^2), standard error of the estimate (S) and Fisher's criterion (F).

Table 1
Rat absorption, Abraham descriptors, ClogP and PSA

No.	Names	Absorption	E	S	A	B	V	ClogP	PSA
	Training set								
1	Caffeine	100 [a]	1.94	1.81	0.00	1.47	1.36	−0.06	47
2	Cimetidine	100 [a]	1.53	2.11	0.59	2.14	1.96	0.35	84
3	Clofibrate	100/100 [a]	0.93	1.23	0.00	0.69	1.82	3.68	31
4	Doxazosin	100 [a]	3.89	4.70	0.25	2.82	3.21	2.37	125
5	Ethinylestradiol	100 [a]	2.12	2.50	0.97	1.16	2.39	3.86	46
6	Miglitol	100 [a]	1.32	1.68	1.20	2.31	1.52	−1.25	100
7	Nimodipine	100 [a]	1.72	2.61	0.32	1.75	3.12	4.14	122
8	Progesterone	100 [a]	1.58	2.47	0.00	1.16	2.62	3.77	30
9	Rimantadine	100	0.88	0.60	0.18	0.65	1.57	2.92	26
10	Salicylicacid	100/100 [a]	1.05	0.89	0.72	0.38	0.99	2.19	55
11	Sultopride	100/100 [a]	1.77	3.25	0.22	2.18	2.71	1.93	68
12	Timolol	100	1.47	1.81	0.10	2.03	2.38	1.61	76
13	Verapamil	100 [a]	1.70	2.48	0.00	2.07	3.79	3.71	64
14	Tinidazole	99	1.40	2.77	0.00	1.36	1.70	−0.50	86
15	Tiopinac	99	2.21	2.11	0.59	0.99	2.03	2.96	52
16	Theophylline	97 [a]	1.93	1.84	0.42	1.38	1.22	−0.06	64
17	Bisoprolol	96 [a]	1.13	1.33	0.10	1.67	2.74	2.12	70
18	Viloxazine	96/100 [a]	1.15	1.42	0.32	1.47	1.87	1.34	45
19	1,3-Diphenyl-1-triazene	95	1.82	1.30	0.20	0.85	1.58	3.99	36
20	Carfecillin	95/95 [a]	2.83	3.31	0.56	2.47	3.20	3.12	111
21	Naproxen	92 [a]	1.62	1.40	0.59	0.75	1.78	2.82	51
22	1,3-Diphenylguanidine	89	2.07	1.84	0.39	1.22	1.72	−0.05	52
23	Dofetilide	88	2.47	3.62	1.10	2.21	3.23	1.58	118
24	Nizatidine	86/100 [a]	1.87	2.55	0.20	2.41	2.46	0.50	83
25	Ramatroban	83	2.58	3.02	1.03	1.61	2.90	3.82	88
26	Tetrapeptide	83	2.92	5.26	1.44	3.58	4.45	1.93	185
27	Casodex	81	2.02	3.85	0.50	1.34	2.71	2.68	92
28	Pentacaine	79	1.64	1.70	0.37	1.40	3.09	5.98	48
29	Saccharin	79/100 [a]	1.59	2.14	0.68	0.95	1.15	0.52	71
30	Dapsone	74	2.32	3.24	0.50	1.25	1.80	1.07	94
31	Acifran	73	1.39	0.88	0.50	1.09	1.56	2.16	68
32	Pelrinone	71 [a]	2.08	2.67	0.61	1.73	1.80	−0.73	90
33	Loprazolam	68	3.60	3.99	0.00	3.05	3.23	NA	91
34	Valaciclovir	68	2.35	2.92	0.59	2.41	2.34	−0.86	147
35	Hydrochlorothiazide	65 [a]	2.19	3.13	1.49	1.78	1.73	−0.40	135
36	Pravastatin	62	1.37	2.08	1.63	1.81	3.37	0.57	112
37	Avitriptan	59	3.11	3.64	0.82	2.96	3.27	2.90	107
38	Colterol	52	1.34	1.15	1.03	1.73	1.84	0.55	77
39	Fosfomycin	48	0.67	1.42	1.52	1.78	0.86	−0.48	79
40	Azithromycin	45 [a]	1.97	3.26	0.93	5.04	6.00	1.83	182
41	Fenoterol	42	2.21	2.16	1.82	2.05	2.36	0.83	105
42	Bumetanide [b]	41/97	2.20	2.73	1.41	1.76	2.64	3.90	121
43	Cyclosporin [b]	40 [a]/100	3.97	6.84	1.54	8.65	10.02	NA	324
44	YM17E	40	3.41	3.49	1.03	2.21	4.57	6.37	79
45	Fosmidomycin	38	0.83	2.44	1.88	2.37	1.23	−3.11	108
46	Doxycycline	33	3.22	3.92	1.62	3.16	3.10	−2.31	162
47	Bromocriptine	32/36 [a]	3.94	4.38	0.84	4.03	4.48	6.69	101
48	Pamaqueside	24	3.29	4.60	1.44	4.92	5.80	−0.20	206
49	Xamoterol	19	1.80	2.37	0.74	2.66	2.57	0.39	109
50	Reproterol	18	3.23	3.15	1.28	3.06	2.81	−0.98	127
51	Acarbose	2	3.31	4.47	2.53	6.19	4.38	−10.62	312
52	Alendronate	0.9	1.15	2.68	3.55	3.59	1.59	−6.22	174
	Test set								
53	Alprenolol	100	1.25	1.03	0.10	1.25	2.16	2.65	43
54	Antipyrine	100 [a]	1.53	1.58	0.00	1.05	1.48	0.41	24
55	Bornaprine(sormodren)	100 [a]	1.29	1.38	0.00	1.20	2.79	4.30	27
56	Cisapride	100/100 [a]	2.30	3.40	0.46	2.04	3.40	3.43	83
57	Codeine	100 [a]	2.02	1.78	0.26	1.75	2.21	0.82	48
58	Diclofenac	100 [a]	1.97	1.88	0.78	0.87	2.03	3.03	40
59	Etintidine	100	1.71	2.25	0.67	2.18	2.15	0.95	83

Table 1 (*Continued*)

No.	Names	Absorption	E	S	A	B	V	ClogP	PSA
60	Felbamate	100	1.44	1.48	0.70	1.12	1.77	−0.29	110
61	Flumazenil	100 [a]	1.80	2.38	0.00	1.76	2.09	1.06	52
62	Fluvastatin	100 [a]	2.39	2.45	1.28	1.60	3.13	4.05	81
63	Imipramine	100 [a]	1.97	1.56	0.00	1.15	2.40	4.41	8
64	Isradipine	100 [a]	1.67	2.46	0.32	1.62	2.71	3.58	95
65	Ketanserin	100 [a]	3.07	3.12	0.33	2.18	2.88	3.00	87
66	Ketoprofen	100 [a]	1.63	1.78	0.59	0.86	1.98	2.76	59
67	Lorcainide	100	2.20	2.31	0.00	1.19	2.96	4.64	19
68	Lormetazepam	100	2.69	2.37	0.10	1.39	2.26	2.60	53
69	Morphine	100 [a]	2.10	1.68	0.55	1.76	2.06	0.24	61
70	Nilvadipine	100	1.67	2.80	0.32	1.69	2.83	2.50	119
71	Oxatomide	100 [a]	3.43	2.83	0.33	2.25	3.40	5.41	44
72	Phenglutamide	100 [a]	1.61	1.89	0.32	1.59	2.39	1.54	49
73	Propranolol	100/99 [a]	1.85	1.36	0.10	1.29	2.15	2.75	43
74	Propylthiouracil	100	1.34	1.40	0.55	1.12	1.28	2.8	44
75	Remoxipride	100 [a]	1.70	2.22	0.22	1.71	2.52	3.26	62
76	Tamsulosin	100 [a]	2.04	2.57	0.47	2.09	3.11	2.17	107
77	Tolmesoxide	100 [a]	1.19	2.21	0.00	1.28	1.62	0.89	37
78	Isoxepac	99 [a]	1.89	2.10	0.59	1.15	1.93	2.57	70
79	Acetylsalicylicacid	98	0.93	1.35	0.59	0.80	1.29	1.02	60
80	Exaprolol	98	1.31	1.10	0.10	1.23	2.52	4.20	37
81	Camazepam	97/97 [a]	2.63	2.56	0.00	1.84	2.67	4.77	52
82	Venlafaxine	97/97 [a]	1.24	1.32	0.35	1.36	2.37	2.11	26
83	Hydrocortisone	95 [a]	2.06	3.16	0.72	1.98	2.80	1.70	96
84	Acetaminophen	92/98 [a]	1.27	1.81	1.02	0.85	1.17	0.49	56
85	Fenclofenac	92/100 [a]	1.80	1.76	0.59	0.62	1.98	4.96	48
86	Granisetron	92/100 [a]	2.18	2.53	0.37	2.03	2.45	1.79	48
87	Nitrendipine	90 [a]	1.72	2.47	0.32	1.53	2.64	4.02	115
88	Torasemide	90	2.14	2.95	1.12	1.90	2.58	3.34	95
89	Enciprazine	89	2.27	2.34	0.17	2.47	3.32	2.76	71
90	Omeprazole	88	2.35	3.41	0.42	2.16	2.52	2.53	72
91	Amlodipine	87	1.82	2.37	0.35	2.07	3.02	3.03	103
92	Ketorolac	87/87 [a]	1.69	2.02	0.59	1.23	1.87	1.62	62
93	Trimethoprim	85	2.52	2.81	0.50	1.76	2.18	0.95	107
94	Bitolterol	82	2.42	2.42	0.19	1.97	3.65	5.39	76
95	Felodipine	81/100 [a]	1.75	2.17	0.32	1.37	2.71	4.53	60
96	Terbutaline	78/60 [a]	1.41	1.40	1.28	1.74	1.84	0.48	80
97	Captopril	> 71 [a]	1.15	1.68	0.50	1.31	1.62	1.19	58
98	Carvedilol	71	3.09	2.49	0.32	2.12	3.10	3.84	80
99	Furosemide	60 [a]	2.05	2.55	1.36	1.47	2.10	1.87	126
100	Atenolol	52/48–50 [a]	1.45	1.89	0.55	1.75	2.18	−0.11	93
101	MK-499	48	2.90	3.54	0.90	2.32	3.44	1.67	109
102	PCE22716	43	2.65	2.18	0.58	2.26	2.50	NA	78
103	Recainam	42	1.26	1.58	0.59	1.33	2.30	1.13	58
104	Ziprasidone	42	3.47	3.04	0.28	1.81	2.92	4.42	57
105	Pamidronicacid	0.5	1.15	2.68	3.55	3.59	1.45	−6.75	173
	Zwitterion								
106	Cefadroxil	95 [a]	2.72	3.07	1.12	2.80	2.49	−2.57	141
107	Levodopa	87 [a]	1.36	1.30	1.36	1.50	1.43	−2.82	114
108	MK-711	78	2.21	1.90	0.59	1.15	2.65	2.55	44
109	Perindopril	67	1.06	1.94	0.28	2.07	2.93	1.34	90
110	Ramipril	56/56 [a]	1.68	2.37	0.28	2.12	3.26	1.67	87
111	Benazepril	50 [a]	2.34	2.43	0.28	1.84	3.27	1.82	85
112	Prulifloxacin	50	2.89	3.48	0.00	2.64	3.01	NA	100
113	Apovincaminc acid	50	2.31	2.00	0.59	1.61	2.44	2.16	48
114	Inogatran	33	1.95	3.13	0.76	3.50	3.51	−0.88	158
115	Enalaprilat	11 [a]	1.60	2.18	0.78	2.08	2.66	0.86	112
116	CGS 16617	7	2.00	2.14	0.96	2.26	2.62	−3.16	137
117	NAD394	5.4	2.61	2.74	0.31	2.26	2.33	1.50	78
118	Amphotericin B	5 [a]	3.70	5.37	3.37	5.70	7.12	NA	296
	Drugs with missing atom values								
119	Clonidine	100 [a]	1.60	1.49	0.64	1.16	1.53	1.37	42

Table 1 (*Continued*)

No.	Names	Absorption	E	S	A	B	V	ClogP	PSA
120	Ximoprofen	100 [a]	1.31	1.34	0.94	0.79	2.07	2.18	77
121	Bretyliumtosylate	20 [a]	0.87	0.62	0.00	0.13	1.72	−1.25	2
122	Azidocillin	41	2.50	2.43	0.40	2.33	2.59	2.50	131
123	Pranolium chloride	18	1.47	1.11	0.26	0.76	2.45	NA	30
124	Oxitropium bromide	14	1.31	1.57	0.35	1.19	2.54	NA	66
125	Ipratropium bromide	17 to 35	1.10	1.27	0.35	0.86	2.73	NA	54
126	Iothalamatesodium	4.2/4 [a]	3.44	3.39	1.57	1.33	2.50	1.42	87
	Less reliable absorption								
127	(−)-6-Aminocarbovir	68	2.16	2.45	1.02	3.04	1.82	NA	125
128	Pidotimod	41	1.59	2.39	0.82	1.95	1.63	−0.22	85
129	Benidipine HCl	58 (48 to 68)	2.66	3.25	0.32	2.24	3.80	5.54	105
130	Sumatriptan	~50/50 [a]	1.87	2.33	0.82	1.88	2.27	0.58	75
131	Sulpiride	43 (35 to 50)	1.84	2.93	0.75	2.21	2.53	1.11	103
132	Carbovir	~40	2.61	2.73	0.92	1.70	1.73	−1.73	111
133	Mespirenone	100	2.57	3.92	0.00	1.86	3.16	2.59	59
134	Nufenoxole	79 (68 to 99)	2.39	2.23	0.00	1.12	3.10	3.49	36
135	L-Canavanine	73	1.04	1.18	0.57	2.28	1.30	NA	151
136	Carbidopa	54	1.49	1.52	1.45	2.08	1.67	−0.44	120
137	Nileprost	44	1.36	2.06	1.28	1.59	3.17	1.75	113
138	ANSA	13	2.43	2.56	0.98	1.36	1.58	−1.00	107
139	Toremifene	80	2.43	2.03	0.02	1.11	3.30	6.35	15
140	Amodiaquine	70	2.73	2.47	0.74	1.87	2.74	5.32	50
141	Prazosin	66	3.40	3.81	0.25	2.38	2.74	2.45	103
142	Crisnatol	64	3.48	2.07	0.29	1.51	2.75	4.03	57
143	Spironolactone	63	2.25	3.74	0.00	1.82	3.17	2.25	60
144	Bromerguride	60	2.58	2.44	0.47	2.04	2.87	3.56	47
145	Lacidipine	~50	1.65	2.34	0.32	1.80	3.62	5.74	67
146	Lovastatin	29/29 [a]	1.29	2.22	0.35	1.32	3.29	4.08	64
147	Idarubicin	~50	3.32	2.52	0.63	2.49	3.47	−0.86	173
148	Nadolol	18/18 [a]	1.61	1.63	0.70	1.88	2.49	0.23	91
149	BMS-18374	12	1.39	2.69	2.04	2.23	3.24	4.20	116
150	Tripeptoid	6	1.78	4.23	0.89	3.15	4.11	3.84	158
151	Diflubenzuron	68 (58 to 77)	1.79	2.67	1.04	0.73	1.99	3.95	55
152	Ranitidine	>63/63 [a]	1.60	2.29	0.20	2.28	2.40	1.33	82
153	Gabapentin (DP) [b]	79 [a]	0.63	0.83	0.77	0.93	1.44	−1.18	66
154	Chlorothiazide (DP) [b]	60 [a]	2.18	3.12	1.21	1.97	1.69	−0.31	128
155	Enalapril (DP) [b]	34 [a]	1.50	2.29	0.28	2.09	2.94	0.79	96
156	Pafenolol (DP) [b]	24 (16 to 31)	1.41	1.75	0.67	2.01	2.84	1.67	86
157	Adefovir (DP) [b]	7.8 [a]	2.08	3.00	1.76	2.72	1.79	−2.08	142
158	Acyclovir (DP) [b]	9/21 [a]	2.34	2.67	0.83	1.87	1.52	−2.07	125

[a] Absorption values obtained from the Chiou and Barve data set [15,16].

[b] Notes for some absorption values: cyclosporin: absorption value of 40% was obtained from Ref. [15]. The absorption of 100% is from the note of Table I of Ref. [15]. Bumetanide: the absorption value of 41% is based on a study using a radioactive method [30]. Absorption of 97% is based on the HPLC method [31]. DP: dose-dependent drugs [15,16,19].

3. Results and discussion

3.1. Relationship between rat intestinal absorption and Abraham descriptors

The linear free energy relationship (LFER) equation (Eq. (2)) developed by Abraham et al. [11,12] is:

$$SP = c + eE + sS + aA + bB + vV \quad (2)$$

Here, SP is a biological or chemical property of a series of compounds in solution, for example the rate of absorption of drugs through a biological membrane. E is the excess molar refraction, S is a combined dipolarity/polarizability descriptor, A is the overall or summation solute hydrogen bond acidity, B is the overall or summation solute hydrogen bond basicity, and V is McGowan's characteristic volume ($cm^3mol^{-1}/100$); c, e, s, a, b, v are equation coefficients, obtained by multiple linear regression. Eq. (2) has recently been applied to human intestinal absorption for 169 drug compounds [8]. There is a significant relationship between the Abraham descriptors and human intestinal absorption; the dominant descriptors in Eq. (2) were hydrogen bond acidity and basicity.

We first applied our QSAR analysis to the Chiou and Barve [15,16] data set for human and rat absorption for 80 organic compounds. Their study showed that there was an excellent linear relationship between human and

rat absorption ($r^2 = 0.97$). To investigate the correlation between Abraham descriptors and rat absorption, we have calculated Abraham descriptors for 78 of these compounds. PEG 900 and PEG 4000, from the Chiou and Barve data set, were removed from this analysis because there are no specific structures for the two drugs. Regression analysis was carried out using the Abraham descriptors and rat absorption for the Chiou and Barve data. However, the relationship (model 1, Table 2) is very poor because the data set contains some charged and zwitterionic compounds for which Abraham descriptors cannot be well calculated at present [20]. Removing these compounds still does not lead to a satisfactory correlation (Fig. 1) ($r^2 = 0.46$, model 2, Table 2) in comparison to human absorption model ($r^2 = 0.83$, see model 8, Table 2). Examining the correlation, we find a number of outliers (bromocriptine, xamoterol, miglitol, lovastatin, acyclovir, nadolol, adefovir, atenolol) in the data set. Because some of these are dose-dependent drugs and absorption values were variable [15,16,19], they were removed from the analysis. Removal of these dose-dependent and zwitterionic compounds improves the regression coefficient to $r^2 =$ 0.77 (model 3, Table 2).

Although a data set of 80 compounds was used by Chiou and Barve [15,16], there are only few compounds available that have low absorption. In addition, there was potentially much more data available from the literature. Therefore, absorption values of 111 compounds were evaluated from original references [19]. Combined with the Chiou and Barve data set, absorption values of 158 drug or drug-like compounds have been collected and are listed in Table 1. The absorption of 32 drugs (drugs 127–158) was considered to be less reliable [15,16,19] and these were removed from the total data set. Although the absorption of 21 drugs (drugs 106–126) was reliable, Abraham descriptors can not be well calculated at present because they are either charged compounds or zwitterions. The Abraham descriptors listed for the compounds in Table 1 are for the neutral species.

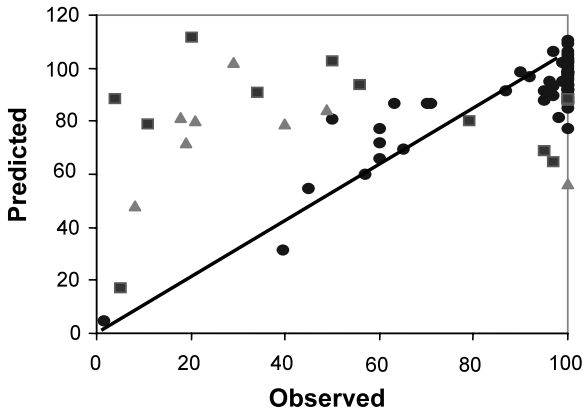

Fig. 1. Plot of observed and predicted % absorption (model 3) for 78 drugs from Chiou's data set (■ zwitterions, ▲ outliers).

The training set of 52 compounds was chosen from compounds 1–105 by use of an alternative space filling design technique developed by Kennard and Stone [21]. The principle of this method is based on the distribution of chosen descriptors, such that descriptors of a training set should cover the descriptor space of the total set. The remaining half number of compounds are used as a test set. The regression analyses obtained for both the rat absorption data set and the previously reported human absorption data [8] are shown in Table 2, models 4 and 8, respectively. The regression results indicate that E has a negative and S has a positive effect on the rat absorption. However, if we check the inter-correlation of descriptors using the regression coefficient (r^2) for 105 drug compounds used, we find that cross-correlation between E and S and between B and V is rather high.

Table 2
Regression results of training and test sets using Abraham descriptors

No.	Data set	Model	Training set					Test sets			
			r^2	n	S	F	RMSE	n	RMSE	AAE	AE
1	Rat (Chiou)	%Abs. = 97 + 4.10E + 1.78S − 23.1A − 7.17B − 1.08V	0.34	78	26	7	25	–	–	–	–
2	Rat (Chiou)	%Abs. = 95 + 6.10E + 1.74S − 23.1A − 15.2B + 5.58V	0.49	66	21	11	20	–	–	–	–
3	Rat (Chiou)	%Abs. = 96 + 6.35E + 2.47S − 20.7A − 15.9B + 6.33V	0.77	58	10	34	11	–	–	–	–
4	Rat	%Abs. = 110 − 13.2E + 10.4S − 25.8A − 9.72B + 1.92V	0.71	52	17	22	16	53	16	12	5
5	Rat	%Abs. = 107 − 25.4A − 6.30B	0.66	52	17	48	15	53	16	11	6
6	Rat (total)	%Abs. = 109 − 23.5A − 7.53B	0.61	105	17	80	16	–	–	–	–
7	Rat-PLS	%Abs. = 109 − 1.47E − 0.22S − 26.6A − 5.26B + 0.17V	0.66	52	17	–	15	53	16	12	4
8	Human	%Abs. = 90 + 2.11E + 1.70S − 20.7A − 22.3B + 15.0V	0.83	38	16	31	14	132	14	11	1
9	Human	%Abs. = 92 − 20.0A − 21.9B + 17.2V	0.82	38	15	53	14	132	14	11	1
10	Human (total)	%Abs. = 96 − 20.0A − 19.8B + 13.9V	0.72	169	15	144	14	–	–	–	–

RMSE: root mean square error, $\mathrm{RMSE} = [\Sigma(\mathrm{Obs} - \mathrm{Calc})^2/n]^{0.5}$. AAE: average absolute error, $\mathrm{AAE} = \Sigma|\mathrm{Obs} - \mathrm{Calc}|/n$. AE: average error, $\mathrm{AE} = \Sigma(\mathrm{Obs} - \mathrm{Calc})/n$. PLS: partial least-squares regression.

	S	A	B	V
E	0.62	0.00	0.32	0.41
S		0.10	0.59	0.55
A			0.24	0.01
B				0.63

Step-wise regression also was carried out between rat absorption and Abraham descriptors. The result shows that the significant descriptors are *A* and *B* (model 5, Table 2). Removing *E*, *S* and *V* from model 4, Table 2 does not significantly affect the regression coefficient and standard error (see model 5, Table 2). PLS was also carried out to find the significant descriptors for same data set. The absorption model with two principal components chosen by PLS (Table 2, model 7) is close to the step-wise regression model (model 5). This result is in agreement with step-wise analysis. The same model 4 could be obtained from PLS if other three components were added in.

Experimental error in obtaining absorption value is quite high, especially for low absorption. Absorption values given from different references could have a 30% difference [19]. This range is reasonable for pharmacokinetic workers but it is difficult for QSAR studies as regards an absorption model. The models in Table 2 both for human and rat have about 15% standard error. If confidence level is 95%, prediction bands will be about 30%; this is considered to be in the range of experimental error. Fig. 2 shows the predicted and observed absorption for the training and the test sets. Most of the predicted absorption values correspond quite well to the observed absorption and the differences are less than 30%; the exceptions being xamoterol (○ in Fig. 2) in the training set and recainam, ziprasidone in the test set (△ in Fig. 2). For compounds with less-reliable absorption data (drugs 127–158), the absorption prediction for some of these drugs is in good agreement with the observed absorption, whilst for others there is little agreement with the observed absorption. For 61% of these less-reliable absorption drugs the difference between predicted and observed absorption is less than 30%. However, if drugs 1–105 are compared (for which the data is considered to be reliable) the percentage of drugs for which the difference between predicted and observed absorption is less than 30% is 92%.

3.2. Comparison between human and rat absorption models

It is useful to examine the similarity between human and rat absorption models and to assess the possible application of the rat model to predict human absorption. We therefore used the rat absorption model 5, Table 2 to predict the human absorption values of 169 drugs. The result (Fig. 3) shows that in general observed human absorption corresponds well to the absorption predicted from the rat absorption model 5, Table 2, especially for high absorption drugs. In comparison with the human absorption predicted from model 9 (Table 2), the rat absorption model was found to predict absorption 8% higher than the human model for 25 low human absorption drugs (1 to 50%) [8] [where, $8\% = \Sigma(\text{Predicted}^{\text{model 5}} - \text{Predicted}^{\text{model 9}})/25$]. For 143 high human absorption drugs (51 to 100%), the predictions from rat absorption model were somewhat low [$\Sigma(\text{Predicted absorption}^{\text{model 5}} - \text{Predicted absorption}^{\text{model 9}})/144 = -5.6\%$]. This general trend can be seen in Fig. 3.

The difference between human and rat absorption models may be due to differences in administration. Nearly all drug administration in humans was by tablet or capsule with 100 to 300 ml of water. In contrast to the human data, all the compounds listed in Table 1 were dosed in solution or suspension, and in many cases with a certain amount of organic solvent such as ethanol, methylcellulose or corn oil. Absorption for drugs dosed in solution is usually equal to or higher (sometimes significantly higher) than in solid forms [22–25].

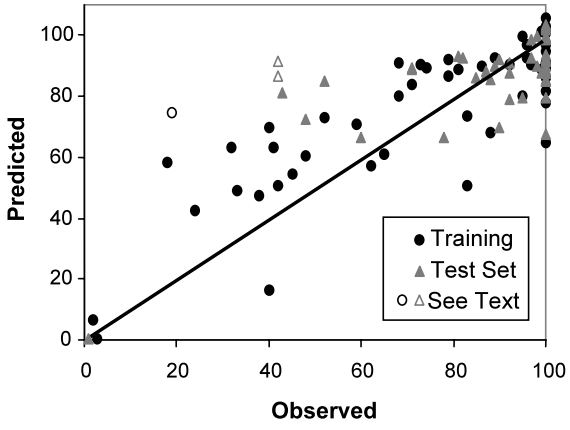

Fig. 2. Plot of predicted (model 4, Table 2) against observed rat % absorption.

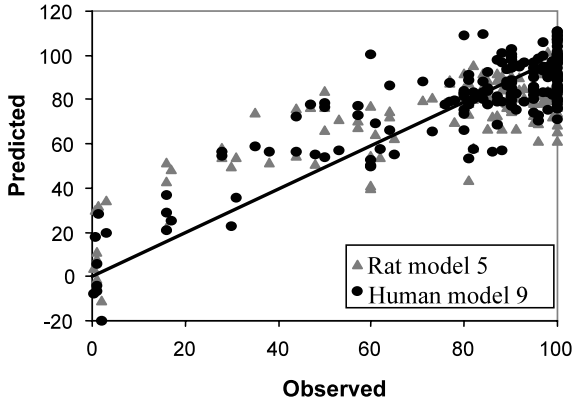

Fig. 3. Prediction of human intestinal % absorption from the rat absorption (model 5, Table 2).

3.3. Non-linear relationship between rat absorption and molecular descriptors

Although the observed absorption in rats is roughly linearly related to the absorption predicted by the Abraham descriptors (Fig. 2), models 5 and 6 in Table 2 predict absorption $>100\%$ for some of the drug compounds and over-predict some poorly absorbed drugs. The same behaviour was found from the linear regression analysis between % absorption and ClogP or PSA (Figs. 4 and 5). Absorption is a kinetic process and molecular descriptors may be linearly related to the absorption kinetic constant, rather than with the percentage of absorption. If diffusion is the rate-limited step [7,8,15] and absorption follow a first-order kinetic process [26–29], % absorption can be converted to a kinetic constant k_{dif} or log k_{dif} from the% absorption by Eq. (3) or Eq. (4) [7,9].

$$\log k_{dif} = \log\left(\ln\frac{1}{1-\%\mathrm{Abs}}\right) - \log t \quad (3)$$

or

$$\%\mathrm{Abs} = 100\times(1-e^{-k_{dif}t}) = 100\times(1-e^{-10^{\log k_{dif}+\log t}}) \quad (4)$$

If we create a set of absorption data from 1 to 99% and use Eq. (3) to calculate the kinetic constant k_{dif}, it can be seen that the relationship between log k_{dif} and % absorption is roughly linear (Fig. 1 in reference 9) but it is not an exact straight line. This explains why regression results are similar using % absorption and logarithm of kinetic constant (see Tables 2 and 3).

Models 1–4 and 9–12 in Table 3 are the regression results between logarithm of the kinetic constant of rat absorption and molecular descriptors by using Eq. (3) for linear and Eq. (4) for non-linear analysis. For linear regression analysis, 0 and 100% absorption values had to be removed from the data set because Eq. (3) collapses when the absorption percent is 0 or 100. As mentioned previously [7,9], the 100% or 0% absorption values observed should be theoretically near 100 or 0, but not equal to 100 or 0 [9]. The advantage of a non-linear model is that the model does not predict absorption $>100\%$ and $<0\%$. In comparing these models, the Abraham method seems better than ClogP and PSA methods. This was also found for linear analyses between percentage of rat absorption and descriptors. Various other non-linear equations did not improve the regression coefficient for ClogP and PSA. Table 3 also lists the regression results between human absorption and descriptors (models 5–8 and 13) published previously for linear and non-linear analysis by using Eqs. (3) and (4), respectively [9].

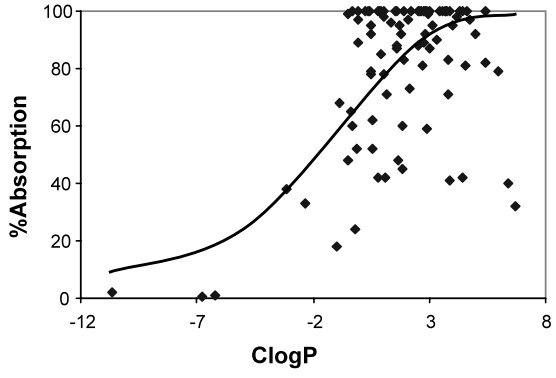

Fig. 4. Plot of rat absorption against ClogP (solid line is model 11).

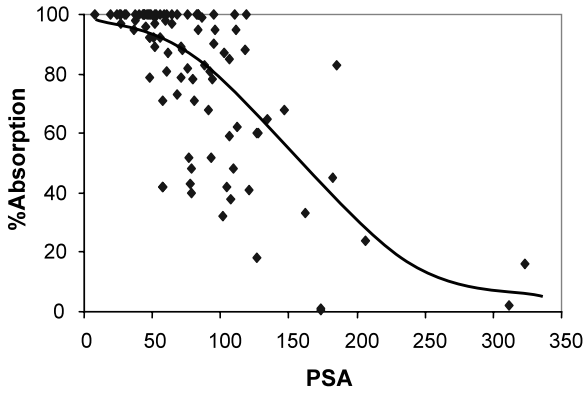

Fig. 5. Plot of rat absorption against polar surface area (solid line is model 12).

4. Conclusions

The LFER equations derived separately to model human and rat absorption are similar but not the same. Hydrogen bond acidity and basicity are the most important descriptors influencing absorption in both humans and rats. To compare the similarity and difference between human and rat models that we established, the rat model was used to predict human intestinal absorption. The results showed that the absorption values predicted from rat model were close to the absorption values predicted from human model for most compounds. However, rat absorption model predicts human absorption 8% higher than the human model for 25 low human absorption drugs. The reason may be due to the difference in oral dosage form administered to the two species. To study the absorption mechanism, percentage of absorption was converted to kinetic constant based on first order kinetic process. Non-linear and linear analysis was carried out between kinetic rate constant and some physicochemical parameters (i.e. octanol/water partition coefficient, polar surface area and Abraham descriptors). The results from non-linear and linear analysis on human and rat absorption are similar.

Table 3
Linear and non-linear analysis between rat and human absorption and molecular descriptors

No.	Method	Model	n	r^2	S
		$\log k_{dif} = \log\{\ln[1/(1-\%Abs)]\} - \log t$			
1	Abraham-rat	$\log k_{dif} = 0.638 - 0.232E + 0.204S - 0.512A - 0.296B + 0.157V$	67	0.79	0.27
2	Abraham-rat-S [a]	$\log k_{dif} = 0.718 - 0.566A - 0.0970B$	67	0.72	0.30
3	ClogP-rat	$\log k_{dif} = -0.140 + 0.128ClogP$	64	0.43	0.44
4	PSA-rat	$\log k_{dif} = 0.658 - 0.00637PSA$	67	0.38	0.44
5	Abraham-human	$\log k_{dif} = 0.568 - 0.036E + 0.141S - 0.413A - 0.507B + 0.232V$	128	0.77	0.31
6	Abraham-human-S [a]	$\log k_{dif} = 0.600 - 0.347A - 0.498B + 0.305V$	128	0.75	0.32
7	ClogP-human	$\log k_{dif} = -0.0532 + 0.195ClogP$	128	0.70	0.35
8	PSA-human	$\log k_{dif} = 0.841 - 0.00835PSA$	128	0.66	0.37
		$\%Abs = 100 \times [1 - EXP(-10^{\log k + \log t})$			
9	Abraham-rat	$\%Abs = 100 \times [1 - EXP(-10^{0.807 - 0.233E + 0.193S - 0.340A - 0.231B + 0.060V})]$	105	0.67	16
10	Abraham-rat-S [a]	$\%Abs = 100 \times [1 - EXP(-10^{0.747 - 0.340A - 0.155B})]$	105	0.62	16
11	ClogP-rat	$\%Abs = 100 \times [1 - EXP(-10^{0.0995 + 0.115ClogP})]$	102	0.35	21
12	PSA-rat	$\%Abs = 100 \times [1 - EXP(-10^{0.688 - 0.00550PSA})]$	105	0.42	20
13	Abraham-human	$\%Abs = 100 \times [1 - EXP(-10^{0.453 + 0.085E + 0.041S - 0.348A - 0.403B + 0.232V})]$	169	0.78	13

[a] Step-wise regression.

References

[1] Y.C. Martin, J. Med. Chem. 24 (1981) 229–237.
[2] J.B. Dressman, G.L. Amidon, D. Fleisher, J. Pharm. Sci. 74 (1985) 588–589.
[3] P. Macheras, M. Symillides, Biopharm. Drug Dispos. 10 (1989) 43–53.
[4] K. Palm, P. Stenberg, K. Luthman, P. Artursson, Pharm. Res. 14 (1997) 568–571.
[5] K. Palm, K. Luthman, A.L. Ungell, G. Strandlund, P. Artursson, J. Pharm. Sci. 85 (1996) 32–39.
[6] D.E. Clark, J. Pharm. Sci. 88 (1999) 807–814.
[7] M.H. Abraham, Y.H. Zhao, J. Le, A. Hersey, C.N. Luscombe, D.P. Reynolds, G. Beck, B. Sherborne, I. Cooper, Eur. J. Med. Chem. 37 (2002) 595–605.
[8] Y.H. Zhao, J. Le, M.H. Abraham, A. Hersey, P.J. Eddershaw, C.N. Luscombe, D. Butina, G. Beck, B. Sherborne, I. Cooper, J.A. Platts, J. Pharm. Sci. 90 (2001) 749–784.
[9] Y.H. Zhao, M.H. Abraham, J. Le, A. Hersey, C.N. Luscombe, G. Beck, B. Sherborne, I. Cooper, Pharm. Res. 19 (2002) 1444–1455.
[10] M.H. Abraham, A. Ibraham, A.M. Zissimos, Y.H. Zhao, J. Comer, D.P. Reynolds, Drug Discov. Today 7 (2002) 1056–1063.
[11] M.H. Abraham, H.S. Chadha, F. Martins, R.C. Mitchell, M.W. Bradbury, J.A. Gratton, Pestic. Sci. 55 (1999) 78–88.
[12] M.H. Abraham, H.S. Chadha, G.S. Whiting, R.C. Mitchell, J. Pharm. Sci. 83 (1994) 1085–1100.
[13] C.A. Lipinski, F. Lombardo, B.W. Dominy, P. Feeney, J. Adv. Drug Deliv. Rev. 23 (1997) 3–25.
[14] M.D. Wessel, P.C. Jurs, J.W. Tolan, S.M. Muskal, J. Chem. Inf. Comput. Sci. 38 (1998) 726–735.
[15] M.L. Chiou, A. Barve, Pharm. Res. 15 (1998) 1792–1795.
[16] W.L. Chiou, H.Y. Jeong, S.M. Chung, T.C. Wu, Pharm. Res. 17 (2000) 135–140.
[17] W.L. Chiou, C. Ma, S.M. Chung, H.Y. Jeong, T.C. Wu, Int. J. Clin. Pharmacol. Ther. 38 (2000) 532–539.
[18] W.L. Chiou, P.W. Buehler, Pharm. Res. 19 (2002) 868–874.
[19] Y.H. Zhao, M.H. Abraham, J. Le, A. Hersey, C.N. Luscombe, G. Beck, B. Sherborne, I. Cooper, Eur. J. Med. Chem. 38 (2003) 233–243.
[20] J.A. Platts, D. Butina, M.H. Abraham, A. Hersey, J. Chem. Inf. Comput. Sci. 39 (1999) 835–845.
[21] R.W. Kennard, L.A. Stone, Technometrics 11 (1969) 137–148.
[22] V. Rodighiero, Clin. Pharmacokinet. 37 (1999) 399–431.
[23] Y.J. Sue, M. Shannon, Clin. Pharmacokinet. 23 (1992) 93–105.
[24] R.D. Faulkner, P. Fernandez, G. Lawrence, L.L. Sia, A.J. Falkowski, A.I. Weiss, A. Yacobi, B.M. Silber, J. Clin. Pharmacol. 28 (1988) 700–706.
[25] G. Flesch, P. Muller, P. Lloyd, Eur. J. Clin Pharmacol. 52 (1997) 115–120.
[26] P.O. Gubbins, K.E. Bertch, Clin. Pharmacokinet. 21 (1991) 431–447.
[27] O.A. Raevsky, V.I. Fetisov, E.P. Trepalina, J.W. McFarland, K.J. Schaper, Quant. Struct.-Act. Relat. 19 (2000) 366–374.
[28] J.B. Dressman, D. Fleisher, J. Pharm. Sci. 75 (1986) 109–116.
[29] M. Rowland, T.N. Tozer, Clinical Pharmacokinetics: Concepts and Applications, Lea & Febiger, Philadelphia, PA, 1995, pp. 113–129.
[30] S.J. Kolis, T.H. Williams, M.A. Schwartz, Drug Metab. Dispos. 4 (1976) 169–176.
[31] S.H. Lee, M.G. Lee, N.D. Kim, J. Pharmacokinet. Biopharm. 22 (1994) 1–17.