articles

molecular pharmaceutics

Prediction and Mechanistic Interpretation of Human Oral Drug Absorption Using MI-QSAR Analysis

Manisha Iyer,[†,‡] Y. J. Tseng,[§,||] C. L. Senese,[§] Jianzhong Liu,[⊥] and A. J. Hopfinger*[,§,⊥]

Laboratory of Molecular Modeling and Design (MC 781), College of Pharmacy, University of Illinois at Chicago, 833 South Wood Street, Chicago, Illinois 60612-7231, Division of Clinical Chemistry, Department of Pathology, Children's Hospital of Pittsburgh, 5834 Main Tower, 200 Lothrop Street, Pittsburgh, Pennsylvania 15213, The Chem21 Group, Inc., 1780 Wilson Drive, Lake Forest, Illinois 60045, Department of Computer Science and Information Engineering, National Taiwan University, No.1 Sec. 4, Roosevelt Road, Taipei, Taiwan 106, and College of Pharmacy, MSC09 5360, 1 University of New Mexico, Albuquerque, New Mexico 87131-0001

Received September 5, 2006; Revised Manuscript Received October 16, 2006; Accepted November 6, 2006

Abstract: Membrane-interaction [MI]-QSAR analysis, which includes descriptors explicitly derived from simulations of solutes [drugs] interacting with phospholipid membrane models, was used to construct QSAR models for human oral intestinal drug absorption. A data set of 188 compounds, which are mainly drugs, was divided into a parent training set of 164 compounds and a test set of 24 compounds. Stable, but not highly fit [$R^2 = 0.68$] MI-QSAR models could be built for all 188 compounds. However, the relatively large number [47] of drugs having 100% absorption, as well as all zwitterionic compounds [11], had to be eliminated from the training set in order to construct a linear five-term oral absorption diffusion model for 106 compounds which was both stable [$R^2 = 0.82$, $Q^2 = 0.79$] and predictive given the test set compounds were predicted with nearly the same average accuracy as the compounds of the training set. Intermolecular membrane−solute descriptors are essential to building good oral absorption models, and these intermolecular descriptors are displaced in model optimizations and intramolecular solute descriptors found in published oral absorption QSAR models. A general form for all of the oral intestinal absorption MI-QSAR models has three classes of descriptors indicative of three thermodynamic processes: (1) solubility and partitioning, (2) membrane−solute interactions, and (3) flexibility of the solute and/or membrane. The intestinal oral absorption MI-QSAR models were compared to MI-QSAR models previously developed for Caco-2 cell permeation and for blood−brain barrier penetration. The MI-QSAR models for all three of these ADME endpoints share several common descriptors, and suggest a common mechanism of transport across all three barriers. A further analysis of these three types of MI-QSAR models has been done to identify descriptor-term differences across these three models, and the corresponding differences in thermodynamic transport behavior of the three barriers.

Keywords: Oral drug absorption; MI-QSAR analysis; membrane barrier transport

Introduction

The oral route for drug delivery has always been strongly preferred over alternative and more invasive routes for systemic administration. Oral drug delivery—specifically tablets, capsules, and soft gels—account for 70% of all dosage

* Corresponding author. Mailing address: College of Pharmacy, MSC09 5360, 1 University of New Mexico, Albuquerque, NM 87131-0001. E-mail: hopfingr@unm.edu. Tel: 505.272.8474. Fax: 505.272.0704.

[†] University of Illinois at Chicago.
[‡] Children's Hospital of Pittsburgh.
[§] The Chem21 Group, Inc.
[||] National Taiwan University.
[⊥] University of New Mexico.

10.1021/mp0600900 CCC: $37.00
Published on Web 12/23/2006

forms consumed.[1] This preference is due to the convenience, relatively low costs, and high patient compliance rates associated with oral drug delivery systems. In an attempt to further enhance patient convenience and compliance by employing flexible dosing conditions, there is now an increased research focus on controlled-release formulations. Consequently, oral dosage forms continue to rise in popularity both among drug developers and among patients, especially those with chronic conditions. Hence, a crucial issue in the development of any new drug is its bioavailablity after oral administration.

The mechanism, prediction, and measurement of human intestinal absorption have been the subject of a number of books and review articles.[2−5] Oral absorption refers to the process of movement of a drug from its site of administration into systemic circulation, while bioavailability is the rate or extent of absorption. Many factors affect the highly complex process of drug absorption, but the three main steps involved are dissolution, diffusion, and perfusion.[3] A solid drug, once administered, needs to first dissolve; the drug in solution then diffuses across the intestinal membrane, and on exit it is removed by perfusion into the blood stream. Thus, drug solubility plays a very important role in absorption. In the case of a poorly soluble drug, dissolution could be the rate-limiting step in the absorption process.[5] On the other hand, for soluble drugs that rapidly diffuse across membrane bilayers of the gastrointestinal tract, perfusion could be the rate-limiting step. Hence, it also stands to reason that for drugs that have diffusion as the rate-limiting step of absorption, dissolution and blood flow will have little effect on their oral bioavailability. A model designed to estimate intestinal drug absorption accounting for all factors involved would be extremely complex.[6] However, based upon the knowledge of the rate-limiting step concerned with the intestinal absorption of a particular drug, various methods can be employed to simplify the procedure. For drugs that are dissolution rate limited, various dissolution tests are used,[5] while for drugs that are diffusion rate limited, animal models like a rat intestinal absorption model[7] or a nonanimal procedure like the Caco-2 monolayer cell model[8] are commonly used.

This paper deals with the application of the membrane-interaction [MI]-QSAR methodology to predict human intestinal absorption of a set of drugs for which diffusion is the rate-limiting step of absorption.

Methods

A. Oral Absorption Data. Zhao et al.[9] collected and evaluated human intestinal absorption data from various literature sources, and this data was divided into diffusion and dissolution rate limited sets of compounds. The set of 188 compounds that have diffusion as the rate-limiting step of absorption has been used in this MI-QSAR study. This set consists of drugs or druglike molecules spanning a wide molecular weight range of 75 to 873 amu and also includes 20 zwitterionic drugs. The oral absorption values of these compounds range from 0.3% to 100%. Table 1 lists the initial training set of 188 compounds with their percentage absorption, molecular weights, [C]logP,[10] and polar surface area, PSA,[11] values.

B. The MI-QSAR Paradigm. 1. Modeling of the Solute Molecules and of the Phospholipid Monolayer. The MI-QSAR paradigm has been discussed in detail previously and is only summarized here.[12−17] Currently, this methodology

(1) Dubin, C. H. Taking Control. *Pharmaceutical Formulation and Quality*, October 2003, 126−144.

(2) Wilson, C. J.: Gastrointestinal transit and drug absorption. In *Oral Drug Absorption: Prediction and Assessment;* Dressman, J. B.; Lennernäs, H., Eds.; Marcel Dekker, Inc.: New York, 2000; pp 1−10.

(3) Abraham, M. H.; Zhao, Y. H.; Le, J.; Hersey, A.; Luscombe, C. N.; Reynolds, D. P.; Beck, G.; Sherborne, B.; Cooper, I. On the mechanism of human intestinal absorption. *Eur. J. Med. Chem.* **2002**, *37*, 595−605.

(4) Binns, T. B. *Absorption and Distribution of drugs*; E. & S. Livingstone LTD: Edinburgh and London, 1964.

(5) Dressman, J. B.; Amidon, G. L.; Reppas, C.; Shah, V. P. Dissolution testing as a prognostic tool for oral drug absorption: immediate dosage forms. *Pharm. Res.* **1998**, *15*, 11−22.

(6) Dressman, J. B.; Amidon, G. L.; Fleisher, D. Absorption potential: estimating the fraction absorbed for orally-administered compounds. *J. Pharm. Sci.* **1985**, *74*, 588−589.

(7) Dolusio, J. T.; Billups, N. F.; Dittert, L. W.; Sugita, E. T.; Swintosky, J. V. Drug absorption I: An *in situ* rat gut technique yielding realistic absorption rates. *J. Pharm. Sci.* **1969**, *58*, 1196−1200.

(8) Artursson, P.; Palm, K.; Luthman, K. Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv. Drug Delivery Rev.* **1996**, *22*, 67−84.

(9) Zhao, Y. H.; Le, J.; Abraham, M. H.; Hersey, A.; Eddershaw, P. J.; Luscombe, C. N.; Butina, D.; Beck, G.; Sherborne, B.; Cooper, I.; Platts, J. A. Evaluation of human intestinal absorption data for use in QSAR studies and a quantative relationship obtained with the Abraham descriptors. *J. Pharm. Sci.* **2001**, *90*, 749−784.

(10) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple method of calculating octanol/water partition coefficient. *Chem. Pharm. Bull. (Tokyo)* **1992**, *40*, 127−130.

(11) Clark, D. E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. *J. Pharm. Sci.* **1999,** *88*, 815−821.

(12) Kulkarni, A. S.; Hopfinger, A. J. Membrane-interaction QSAR analysis: application to the estimation of eye irritation by organic compounds. *Pharm. Res.* **1999**, *16*, 1244−1252.

(13) Kulkarni, A. S.; Hopfinger, A. J.; Osborne, R.; Bruner, L. H.; Thompson, E. D. Prediction of Eye Irritation from Orgranic Chemicals Using Membrane-Interaction QSAR Analysis. *Toxicol. Sci.* **2001,** *59*, 335−345.

(14) *ADMET/MI-QSAR Version 1.0 User's Guide*; The Chem21 Group, Inc.: Chicago, 2001.

(15) Kulkarni, A.; Han, Y.; Hopfinger, A. J. Predicting Caco-2 cell permeation coefficients of organic molecules using membrane-interaction QSAR analysis. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 331−342.

(16) Iyer, M.; Mishra, R.; Han, Y.; Hopfinger, A. J. Predicting blood-brain barrier partitioning of organic molecules using membrane-interaction QSAR analysis. *Pharm. Res.* **2002**, *19*, 1611−1621.

Table 1. Percentage of Absorption, Molecular Weight, Octanol−Water Partition Coefficient, and Polar Surface Area of the Complete Data Set Which Is Training Set A for Constructing Eqs 6 and 7

no.	drug name	% Abs	MW (amu)	ClogP	PSA	no.	drug name	% Abs	MW (amu)	ClogP	PSA
1	aminopyrine	100	231.30	1.0	25.0	70	sotalol	95	272.36	0.2	85.0
2	bornaprine	100	253.38	4.3	27.0	71	timolol	95	316.42	1.6	76.0
3	caffeine	100	194.20	−0.1	47.0	72	alprenolol	93	249.34	2.7	43.0
4	camazepam	100	371.81	3.6	52.0	73	amrinone	93	187.20	−0.6	75.0
5	cicaprost	100	374.46	2.0	99.0	74	isradipine	92	371.39	3.6	95.0
6	cisapride	100	465.94	3.4	83.0	75	ketoprofen	92	254.27	2.8	59.0
7	corticosterone	100	346.45	2.3	73.0	76	hydrocortisone	91	362.45	1.7	96.0
8	cyproterone acetate	100	416.92	3.4	49.0	77	naloxone	91	327.37	0.0	69.0
9	desipramine	100	266.38	4.1	20.0	78	alprazolam	90	308.76	2.3	39.0
10	diazepam	100	284.73	3.3	28.0	79	amphetamine	90	135.20	1.6	27.0
11	diclofenac	100	296.14	3.0	40.0	80	betaxolol	90	307.42	2.2	55.0
12	ethinyl estradiol	100	296.39	3.7	46.0	81	chloramphenicol	90	323.13	0.7	118.0
13	fenclofenac	100	297.12	5.0	48.0	82	felbamate	90	238.24	−0.3	110.0
14	fluvastatin	100	411.46	3.2	76.0	83	ketorolac	90	255.26	1.6	62.0
15	gallopamil	100	484.62	3.1	68.0	84	meloxicam	90	351.39	3.1	101.0
16	glyburide	100	493.99	4.1	110.0	85	nisoldipine	90	388.41	4.2	82.0
17	granisetron	100	312.41	1.8	48.0	86	nizatidine	90	331.46	0.5	83.0
18	imipramine	100	280.40	4.4	8.0	87	phenytoin	90	252.27	2.1	59.0
19	indomethacin	100	357.78	4.2	68.0	88	sulindac	90	356.40	2.8	58.0
20	isoxicam	100	335.33	2.4	116.0	89	terazosin	90	387.44	2.7	102.0
21	levonorgestrel	100	312.43	3.3	40.0	90	tramadol	90	263.37	2.3	22.0
22	lormetazepam	100	335.18	2.6	53.0	91	dihydrocodeine	89	301.37	1.3	49.0
23	lornoxicam	100	371.81	3.2	100.0	92	oxazepam	89	286.71	2.3	67.0
24	mexiletine	100	179.26	2.6	34.0	93	sultopride	89	354.46	1.9	68.0
25	nefazodone	100	470.01	5.0	51.0	94	tenidap	89	320.74	0.6	77.0
26	nicotine	100	162.23	1.3	15.0	95	felodipine	88	384.24	5.0	60.0
27	ondansetron	100	293.36	2.6	31.0	96	moxonidine	88	241.69	1.0	69.0
28	oxatomide	100	426.55	5.4	44.0	97	nitrendipine	88	360.36	3.4	105.0
29	phenglutarimide	100	288.38	1.5	49.0	98	saccharin	88	183.18	0.5	71.0
30	piroxicam	100	331.34	2.7	99.0	99	bupropion	87	239.73	3.2	24.0
31	praziquantel	100	312.40	3.4	36.0	100	lamivudine	87	229.26	−1.5	93.0
32	progesterone	100	328.48	3.8	30.0	101	pindolol	87	248.32	1.7	63.0
33	salicylic acid	100	138.12	2.2	55.0	102	topiramate	86	339.36	−0.1	121.0
34	stavudine	100	224.22	−0.5	86.0	103	lansoprazole	85	349.31	3.1	65.0
35	sudoxicam	100	337.37	2.6	101.0	104	morphine	85	285.33	0.2	61.0
36	tenoxicam	100	337.37	2.4	100.0	105	oxyfedrine	85	313.38	2.8	57.0
37	testosterone	100	288.41	3.2	40.0	106	tolbutamide	85	270.34	2.5	78.0
38	theophylline	100	180.17	−0.1	64.0	107	aspirin	84	180.15	1.0	60.0
39	toremifene	100	405.94	6.4	15.0	108	bromazepam	84	316.15	1.7	53.0
40	valproic acid	100	144.21	2.8	40.0	109	captopril	84	217.28	1.2	58.0
41	verapamil	100	454.59	3.7	64.0	110	propiverine	84	367.47	4.1	28.0
42	carfecillin	99	454.49	3.1	111.0	111	methylprednisolone	82	374.46	2.0	95.0
43	naproxen	99	230.25	2.8	51.0	112	mifobate	82	358.64	0.7	70.0
44	nordiazepam	99	270.71	3.0	43.0	113	sorivudine	82	349.13	−1.7	127.0
45	prednisolone	99	360.43	1.6	97.0	114	digoxin	81	780.92	1.3	216.0
46	propranolol	99	259.34	2.8	43.0	115	flecainide	81	414.35	4.4	55.0
47	atropine	98	289.36	1.3	50.0	116	piroximone	81	217.23	1.0	82.0
48	lamotrigine	98	256.10	3.2	97.0	117	quinidine	81	324.41	2.9	40.0
49	minoxidilne	98	210.27	1.1	94.0	118	acebutolol	80	336.42	1.6	88.0
50	tolmesoxide	98	214.27	0.9	37.0	119	acetaminophen	80	151.16	0.5	56.0
51	viloxazine	98	237.29	1.3	45.0	120	dexamethasone	80	392.45	2.0	90.0
52	warfarin	98	308.32	2.4	51.0	121	ethambutol	80	204.31	0.1	69.0
53	antipyrine	97	188.23	0.4	24.0	122	guanabenz	80	231.08	3.0	76.0
54	clofibrate	97	242.69	3.7	31.0	123	isoniazid	80	137.15	−0.7	72.0
55	disulfiram	97	296.52	3.9	5.0	124	methadone	80	309.44	3.1	16.0
56	trimethoprim	97	290.32	1.0	107.0	125	omeprazole	80	345.41	2.5	72.0
57	venlafaxine	97	277.40	2.1	26.0	126	urapidil	78	387.48	2.6	65.0
58	bumetanide	96	364.41	3.9	121.0	127	famciclovir	77	321.34	−0.4	113.0
59	torasemide	96	348.42	3.3	95.0	128	mercaptoethanesulfonic acid	77	142.19	−0.5	59.0
60	trapidil	96	205.27	1.9	43.0	129	propylthiouracil	76	170.23	2.8	44.0
61	codeine	95	299.36	0.8	48.0	130	cycloserine	73	102.10	−1.7	80.0
62	fluconazole	95	306.29	−0.1	61.0	131	recainam	71	263.38	1.1	58.0
63	flumazenil	95	289.27	1.1	52.0	132	hydrochlorothiazide	69	297.73	−0.4	135.0
64	ibuprofen	95	206.27	3.7	40.0	133	cimetidine	64	252.35	0.4	84.0
65	labetolol	95	328.40	2.5	95.0	134	metolazone	64	365.83	2.4	96.0
66	metoprolol	95	267.36	1.2	56.0	135	terbutaline	62	225.28	0.5	80.0
67	oxprenolol	95	265.34	1.7	53.0	136	furosemide	61	330.74	1.9	126.0
68	practolol	95	266.34	0.8	77.0	137	fenoterol	60	303.35	0.8	105.0
69	scopol	95	303.35	0.3	61.0	138	pirbuterol	60	240.30	−0.9	90.0

Table 1 (Continued)

no.	drug name	% Abs	MW (amu)	ClogP	PSA	no.	drug name	% Abs	MW (amu)	ClogP	PSA
139	reproterol	60	389.41	−1.0	127.0	154	fosmidomycin	30	183.10	−3.1	108.0
140	zipasidone	60	412.93	4.4	57.0	155	lincomycin	28	406.53	−0.1	125.0
141	nadolol	57	309.40	0.2	91.0	156	netivudine	28	282.25	−2.0	131.0
142	Sumatriptan	57	295.40	0.6	75.0	157	adefovir	16	274.20	−2.1	142.0
143	metformin	53	129.18	−2.6	86.0	158	k-strophanthoside	16	872.93	−5.4	273.0
144	amiloride	50	229.64	−0.3	157.0	159	mannitol	16	182.17	−4.7	129.0
145	atenolol	50	266.34	−0.1	93.0	160	cidofovir	3	279.19	−3.6	156.0
146	guanoxan	50	207.23	0.3	87.0	161	ganciclovir	3	255.24	−3.0	146.0
147	rimiterol	48	223.27	0.4	79.0	162	acarbose	2	645.60	−10.6	321.0
148	cymarin	47	548.65	−0.2	126.0	163	ouabain	1.4	584.64	−4.6	196.0
149	metaproterenol	44	211.26	0.1	81.0	164	kanamycin	1	484.51	−7.8	295.0
150	sulpiride	44	341.42	1.1	103.0	165	neomycin	1	614.66	−9.0	354.0
151	famotidine	38	337.45	−0.6	182.0	166	streptomycin	1	581.59	−7.2	346.0
152	ascorbic acid	35	176.12	−2.2	120.0	167	lactulose	0.6	342.30	−5.6	208.0
153	fosfomycin	31	138.06	−0.5	79.0	168	raffinose	0.3	504.44	−8.0	288.0
						Zwitterions[a]					
169	cefadroxil	100	363.39	−2.6	141.0	179	nicotinic acid	88	123.11	0.8	50.0
170	cephalexin	100	347.39	−1.9	117.0	180	trovaflaxicin	88	416.36	−1.2	97.0
171	glycine	100	75.07	−3.2	73.0	181	levodopa	86	197.19	−2.8	114.0
172	loracarbef	100	349.77	−0.5	117.0	182	cefatrizine	75	462.50	−3.0	184.0
173	ofloxacin	100	361.37	−0.2	73.0	183	ampicillin	62	349.40	−1.3	116.0
174	pefloxacin	100	333.36	0.1	63.0	184	vigabatrin	58	129.16	−2.9	69.0
175	amoxicillin	93	365.40	−1.9	140.0	185	eflornithine	55	182.18	−3.0	94.0
176	telmisartan	90	512.63	7.3	63.0	186	tranexamic acid	55	157.21	−1.8	70.0
177	tiagabine	90	375.53	2.8	45.0	187	methyldopa	41	211.21	−2.1	109.0
178	acrivastine	88	348.43	1.1	53.0	188	ceftriaxone	1	554.58	−2.1	212.0

[a] The definition of zwitterionic compounds is based on the presence of both an ionizable acid group (carboxylic acid or a hydrogen-bearing tetrazole) and an ionizable base group (primary, secondary, tertiary amine or a pyridine). These compounds may or may not be zwitterions according to their pK_a values.

uses a model membrane monolayer composed of dimyristoylphosphatidylcholine (DMPC) molecules. DMPC is modeled from available crystal structure data.[18] The structure of a DMPC molecule is shown in Figure 1. An assembly of 25 DMPC molecules ($5 \times 5 \times 1$) in (x,y,z) directions, respectively, is the model membrane monolayer (Figure 2). Additional information regarding the construction of the monolayer model can be found in refs 12−17.

The DMPC molecule, the training set, and the test compounds [Table 1] were built using the HyperChem program,[19] and the AM1 Hamiltonian in Mopac 6.0[20] was used for the estimation of partial atomic charge distributions over the molecules.

2. Molecular Dynamic Simulations, MDS. The conditions set for the MDS were established in previous MI-QSAR analyses[12−17] and are only summarized here. An initial MDS on the model membrane, without a solute molecule present, was carried out to allow structural relaxation and distribution of the kinetic energy over the monolayer. In order to prevent unfavorable van der Waals interactions between a solute molecule and the membrane DMPC molecules, one of the "center" DMPC molecules was removed from the equilibrated monolayer and a test solute molecule inserted in the space created by the missing DMPC molecule. Each of the test solute molecules of the permeation data set was inserted at three different positions (depths) in the DMPC monolayer with the most polar group of the solute molecule "facing" toward the head group region of the monolayer. Three corresponding MDS models were generated for each solute molecule with regard to the trial positions of the solute molecule in the monolayer. The three trial positions were

1. solute molecule in the head group region
2. solute molecule in between the head group region and the aliphatic chains
3. solute molecule in the tail region of the aliphatic chains

The lowest energy geometry of the solute molecule in the monolayer was sought using each of the three trial solute positions. The three initial MDS positions of ethanol are shown in Figure 3a to illustrate this modeling procedure. The energetically most favorable geometry of this solute molecule in the model DMPC monolayer from all three MDS is shown in Figure 3b.

MDS were carried out using the Molsim package with an extended MM2 force field.[21] The simulation temperature was

(17) Santos-Filho, O. A.; Hopfinger, A. J.; Zheng, T. Characterization of skin penetration processes of organic molecules using molecular similarity and QSAR analysis. *Mol. Pharm.* **2004**, *1*, 466−476.
(18) Hauser, H.; Pascher, I.; Pearson, R. H.; Sundell, S. Preferred conformation and molecular packing of phosphatidylethanolamine and phosphatidylcholine. *Biochim. Biophys. Acta* **1981**, *650*, 21−51.
(19) *Hyperchem Release 4.5 for MS Windows*; Hypercube Inc: Waterloo, Ontario, 1998.
(20) *Mopac 6.0*; Frank J. Seiler Research laboratory, United States Air Force Academy: 1990.
(21) Doherty, D. C. *Molsim Version 3.0 User's Guide*; The Chem21 Group, Inc.: 1780 Wilson Drive, Lake Forest, IL, 2000.

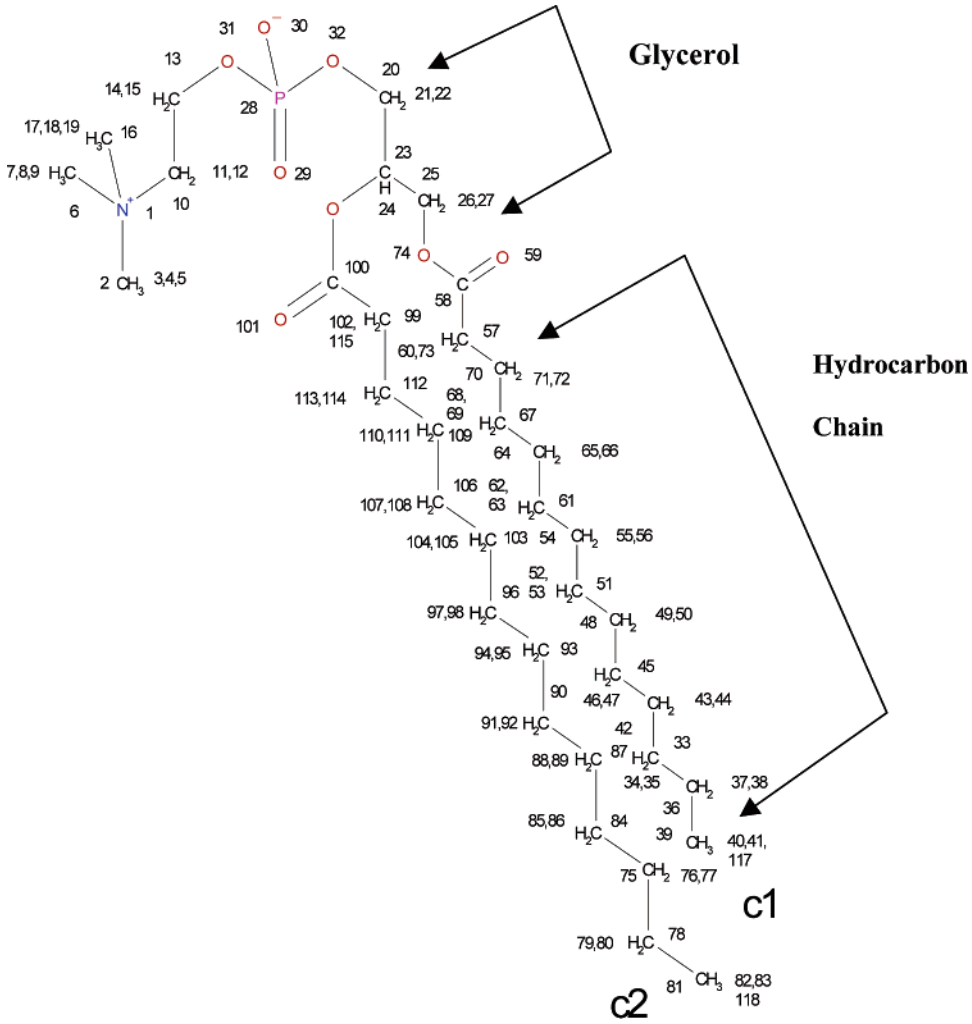

Figure 1. The chemical structure of a DMPC phospholipid molecule with an arbitrary atom numbering assignment. **c1** and **c2** denote the two aliphatic chains of a DMPC molecule.

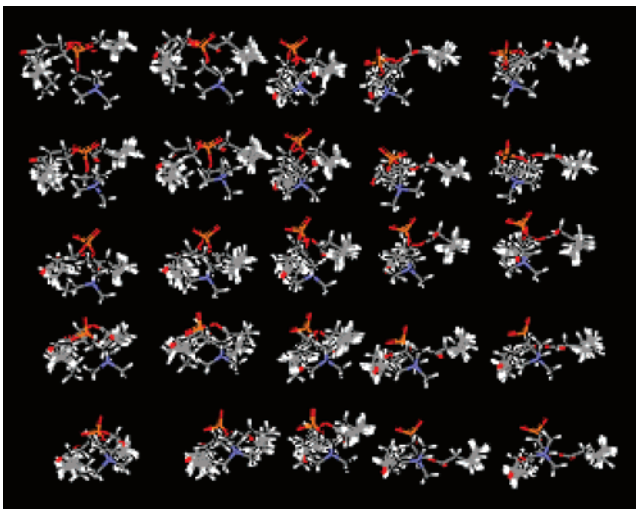

Figure 2. Top view of the monolayer assembly.

set at 311 K, and was held constant in the MDS by coupling the system to an external fixed temperature bath.[22] The trajectory step size was 0.001 ps over a total simulation time of 20 ps for each test compound. The system was sampled every 100 steps to yield a sampling set of 2,000 states. Two-dimensional periodic boundary conditions corresponding to x and y sides of model membrane, but not the "surface plane" of the monolayer, were employed ($a = 50$ Å,^{2} $b = 50$ Å ,^{2} $c = 80$ Å^2 and $\gamma = 90°$) for the DMPC molecules of the monolayer model, but not the test solute molecule. The angle γ is the angle an extended conformation DMPC molecule makes with the "planar surface" of the monolayer.

Only a single solute molecule was explicitly considered in each MDS. Each of the solute molecules, at the start of an MDS, was placed at each of the three different positions in the monolayer, as described above, with the most polar portion of the solute "facing" toward the head group region.

3. Calculation of Descriptors and the Construction of MI-QSAR Models. The descriptors used in the MI-QSAR analysis can be divided into (a) *general intramolecular solute* descriptors, (b) *solute−membrane intermolecular* descriptors, and (c) *solute aqueous dissolution and solvation* descriptors.

(22) Berendsen, H. J. C.; Postman, J. P. M.; Gunsteren, W. F. v.; Nola, A. D.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

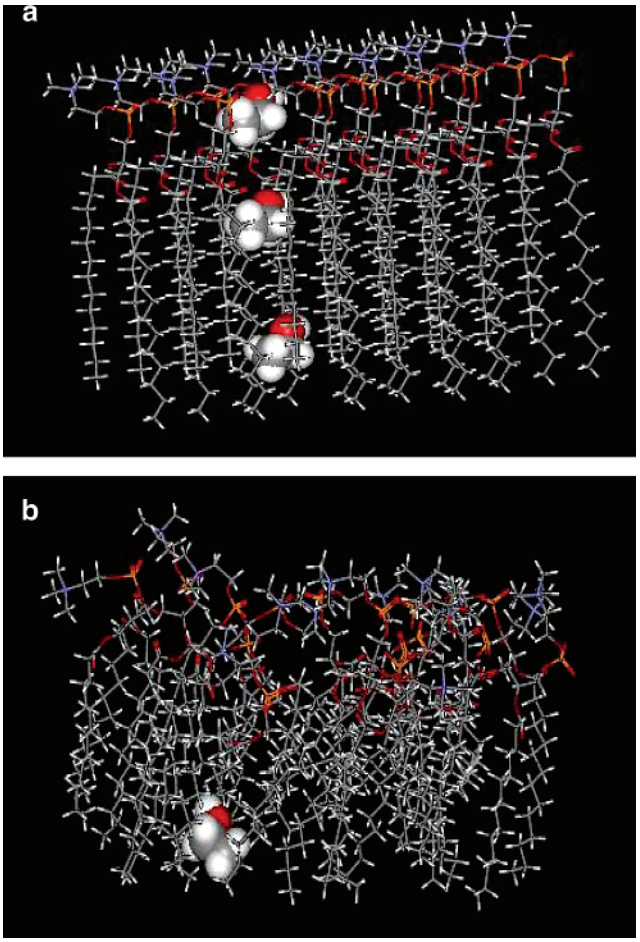

Figure 3. (a) A "side" view of an ethanol molecule inserted at three different positions in the DMPC model monolayer prior to the start of each of the three corresponding MDS used in the MI-QSAR modeling. (b) The lowest energy geometry of a DMPC−ethanol complex in the MDS.

The *general intramolecular solute descriptors* included as part of the trial descriptor pool are listed and defined in Table 2. It is to be noted that the ClogP (Table 1) was calculated using Daylight software,[23] and the PSA values (Table 1) of both the training and test set molecules were taken from the study reported by Zhao et al.[24]

The *intermolecular solute−membrane interaction descriptors* are extracted directly from the MDS trajectories and are listed in part A of Table 3. These particular intermolecular descriptors are calculated using the most stable (lowest total potential energy) solute−membrane geometry realized from MDS sampling of the three initial positions (see Figure 3a) for each of the solutes.

It should be noted that $F(H_2O)$, F(oct), and LogP, the aqueous and 1-octanol solvation free energies of the solutes and the corresponding 1-octanol/water partition coefficient, respectively, are computed using intramolecular computational methods. This is also true for E(coh), T_M, and T_G, the cohesive energy and the hypothetical crystal-melt and glass transition temperatures of the solutes, respectively, which are used to estimate solute dissolution properties. However, all six of these descriptors are intermolecular properties, the first three relating to solute solvation, and the last three to solute dissolution. Therefore, these descriptors are classified as *solvation and dissolution intermolecular descriptors* and listed in part B of Table 3.

Table 2. The General Intramolecular Solute Descriptors Used in the Trial MI-QSAR Descriptor Pool

HOMO	highest occupied molecular orbital energy
LUMO	lowest unoccupied molecular orbital energy
Dp	dipole moment
Vm	molecular volume
SA	molecular surface area
Ds	density
MW	molecular weight
MR	molecular refractivity
N(hba)	number of hydrogen bond acceptors
N(hbd)	number of hydrogen bond donors
N(B)	number of rotatable bonds
JSSA (X)	Jurs−Stanton surface area descriptors
Chi-N, Kappa-M	Kier and Hall topological descriptors
Rg	radius of gyration
PM	principal moment of inertia
PSA	polar surface area
Se	conformational entropy
Q(I)	partial atomic charge densities

4. Construction and Testing of Intestinal Absorption MI-QSAR Models. All MI-QSAR models reported in this study are built using multidimensional linear regression fitting, and the models are optimized by employing the genetic function approximation (GFA). GFA is a multidimensional optimization method based on the genetic algorithm paradigm.[25] Both linear and quadratic representations of each of the descriptor values are included in the trial descriptor pool, and MI-QSAR models are built as a function of number of descriptor terms in a model. Statistical significance in the optimization of an MI-QSAR model is judged using both the correlation coefficient of fit, R^2, and the leave-one-out (LOO) cross-validation correlation coefficient, Q^2. In addition, random scrambling of the dependent variable [20 randomly generated data sets from each training set] is carried out, and an attempt is made to construct corresponding statistically significant MI-QSAR models. No statistically significant randomly scrambled MI-QSAR models were found for any of the data sets investigated in this study. Covariance among the significant descriptors in the optimized MI-QSAR models is evaluated by constructing

(23) . *ClogP Daylight Chemical Information Software, version 4.51*; Daylight Chemical Information Inc.: Los Altos, CA, 1998.

(24) Zhao, Y. H.; Abraham, M. H.; Le, J.; Hersey, A.; Luscombe, C. N.; Beck, G.; Sherborne, B.; Cooper, I. Rate-Limited steps of human oral absorption and QSAR studies. *Pharm. Res.* **2002**, *19*, 1446−1457.

(25) Rogers, D.; Hopfinger, A. J. Applications of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.

Table 3. The Intermolecular Interaction Descriptors in the Trial Descriptor Pool

	A. Solute−Membrane Intermolecular Descriptors
$<F(\text{total})>$	average total free energy of interaction of the solute and membrane
$<E(\text{total})>$	average total interaction energy of the solute and membrane
$E_{\text{INTER}}(\text{total})$	interaction energy between the solute and the membrane at the total intermolecular system minimum potential energy
$E_{XY}(Z)_E$	Z = 1,4-nonbonded, general van der Waals, electrostatic, hydrogen bonding, torsion, and combinations thereof energies at the total intermolecular system minimum potential energy.
	X, Y can be the solute, S, and/or membrane, M, and if E = free, then X = Y = S and the energies are for the solute not in the membrane, but isolated by itself.
$\Delta E_{XY}(Z)$	Change in the Z = 1,4-nonbonded, general van der Waals, electrostatic, hydrogen bonding, torsion, and combinations thereof energies due to the uptake of the solute to the total intermolecular system minimum potential energy.
	X, Y can be the solute, S, and/or membrane, M.
$E_{TT}(Z)$	Z = 1,4-nonbonded, general van der Waals, electrostatic, hydrogen bonding, torsion, and combinations thereof energies of the total [solute and membrane model] intermolecular minimum potential energy.
$\Delta E_{TT}(Z)$	change in the Z = 1,4-nonbonded, general van der Waals, electrostatic, hydrogen bonding, and combinations thereof of the total [solute and membrane model] intermolecular minimum potential energy
ΔS	change in entropy of the membrane due to the uptake of the solute
S	absolute entropy of the solute−membrane system
$\Delta\rho$	change in density of the model membrane due to the permeating solute
$<d>$	average depth of the solute molecule from the membrane surface
	B. Solute Aqueous Dissolution and Solvation Descriptors
$F(H_2O)$	aqueous solvation free energy
F(oct)	1-octanol solvation free energy
logP	1-octanol/water partition coefficient
E(coh)	cohesive packing energy of the solute molecules
T_M	hypothetical crystal-melt transition temperature of the solute
T_G	hypothetical glass transition temperature of the solute

the linear cross-correlation matrix of the descriptors, and by comparing relative descriptor usage in the crossover optimization process of the GFA analysis.

C. The Diffusion Rate Constant. For an aqueous soluble drug, its permeability characteristics play a major role in its absorption across the gastrointestinal membrane. Passive diffusion can be generally described by Fick's law,[24,26] according to which the rate of diffusion is a function of the concentration gradient, the surface area and distance (thickness of the membrane) involved, and characteristic physicochemical properties of the biological barrier and the diffusing substance. There is usually a sufficient quantity of a soluble drug dissolved in the small intestinal fluid so that the drug concentration on the receiving site (portal vein) is often negligible in comparison. Consequently, the rate-determining step for absorption is the passive diffusion through the membrane and the percentage of absorption is directly related to the diffusion rate.[27] If the rate of diffusion follows first-order kinetics,[28,29] then the percentage of absorption (% Abs), or fraction absorbed (FA), and the diffusion rate constant (k_{dif}) are related as given in the following equations: since

$$\mathrm{d}C_{\text{I}}/\mathrm{d}t = -k_{\text{dif}}C_{\text{I}} \tag{1}$$

$$\ln(C_{\text{I}}^{\,0} - C_{\text{p}}^{\,t})/C_{\text{I}}^{\,0} = -k_{\text{dif}}t \tag{2a}$$

$$C_{\text{p}}^{\,t}/C_{\text{I}}^{\,0} = \text{FA} \tag{2b}$$

$$\ln(1 - \text{FA}) = -k_{\text{dif}}t \tag{3}$$

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

$$\log[\ln(1/1 - \text{FA})] = \log k_{\text{dif}} + \log t \tag{5}$$

In eqs 1−5, $\mathrm{d}C_{\text{I}}/\mathrm{d}t$ is the diffusion rate through the gastrointestinal membrane, k_{dif} is the diffusion rate constant, C_{I} is the drug concentration in the intestinal fluid, $C_{\text{I}}^{\,0}$ is the initial concentration in the intestinal fluid, $C_{\text{p}}^{\,t}$ is the concentration in the portal vein at time t, and log t is a constant when it is assumed that the transit time is the same across the gastrointestinal tract for all drugs. The implications of such an assumption are discussed in more detail below.

In this MI-QSAR study, both % Abs and log k_{dif} are used as dependent variables to construct human oral absorption MI-QSAR models.

Results

The two best MI-QSAR models, eqs 6 and 7, for the initial data set comprising all 188 drug molecules (Table 1) are presented in Table 4 along with their R^2 (correlation

(26) Washington, N.; Washington, C.; Wilson, C. G. *Physiological Pharmaceutics, Barriers to drug Absorption*, 2nd ed.; Taylor and Francis, London, 2001.
(27) Martin, Y. C.; Kutter, E.; Austel, V.: *Modern Drug Research—Paths to Better and Safer Drugs*; Dekker: New York, 1989.
(28) Smith, D. A.; van de Waterbeemd, H.; Walker, D. K. *Pharmacokinetics and Metabolism in Drug Design*; Wiley-VCH: Weinheim, New York, 2001.
(29) Rowland, M.; Tozer, T. N. *Clinical Pharmacokinetics: Concepts and Applications*; Lea & Febiger: Philadelphia, 1989.

Table 4. Percentage Oral Absorption, % Abs, MI-QSAR Models for the Initial Training Set (A)

eq	terms	*N*	model	R^2	Q^2
6	7	188	% Abs = 78.32 + 0.13 ΔE_{TT} (hb) + 3.39 ClogP − 0.03 ΔE_{TT}(total) + 0.31$F(H_2O)$ + 0.05$E_{SS}(1-4)_{free}$ + 0.04T_G − 39.24 Dp	0.68	0.65
7	6	188	% Abs = −10.05 + 0.05$E_{TT}(1-4)$ − 0.13 PSA + 0.04 T_G + 0.12 ΔE_{TT}(hb) + 3.62 ClogP − 0.06E_{TT}(vdw)	0.67	0.64

Table 5. Cross-Correlation Matrix of Percentage of Absorption of the MI-QSAR Descriptors of Eqs 6 and 7[a]

	ClogP	PSA	$F(H_2O)$	T_G	Dp	$E_{SS}(1-4)_{free}$	$E_{TT}(1-4)$	E_{TT}(vdw)	ΔE_{TT}(hb)	ΔE_{TT}(total)
ClogP	1.000									
PSA	0.577	1.000								
$F(H_2O)$	0.542	**0.863**	1.000							
T_G	0.016	0.002	0.022	1.000						
Dp	0.012	0.040	0.003	0.099	1.000					
$E_{SS}(1-4)_{free}$	0.001	0.001	0.018	0.013	0.003	1.000				
$E_{TT}(1-4)$	0.000	0.000	0.012	0.007	0.007	**0.887**	1.000			
E_{TT}(vdw)	0.003	0.001	0.000	0.010	0.005	0.000	0.000	1.000		
ΔE_{TT}(hb)	0.500	**0.639**	0.585	0.000	0.002	0.013	0.017	0.000	1.000	
ΔE_{TT}(total)	0.005	0.000	0.000	0.000	0.000	0.000	0.001	0.380	0.001	1.000

[a] Highly correlated descriptors are shown in bold.

coefficient of determination) and Q^2 (cross-validated coefficient of determination) values. One of the advantages of performing GFA model optimization is the generation of multiple significant models, as opposed to a single model generated by other model optimization methods. Both MI-QSAR models, eqs 6 and 7, have a number of descriptors in common, and very similar R^2 and Q^2 values. It was found as part of the GFA optimization process that models with more than seven terms tend to be overfit as indicated by a drop in their Q^2 values compared to corresponding six- and seven-term models.

Other than the partition coefficient, ClogP, the descriptors that are common to both models are ΔE_{TT}(hb), which is the change in the total hydrogen-bonding energy upon uptake of the solute (drug) molecule into the DMPC membrane system, and T_G, which is the hypothetical glass transition temperature of the solute molecule, and models the dissolution of a liquid or gel-like solute. Both of these descriptors are highly indicative of the flexibility [conformational entropy] of a molecule, and/or a molecular complex like the membrane−solute system. As overall hydrogen bonding is lost upon uptake of a solute into a membrane, molecular flexibility of the complex increases. As the structural groups composing a polymer becomes more rigid, its T_G generally increases and the molecular flexibility of the polymer decreases. The positive regression coefficients for both ΔE_{TT}(hb) and T_G in eqs 6 and 7 indicate that % Abs increases as molecular flexibility decreases. Decreasing molecular flexibility corresponds to decreasing favorable solute−membrane binding interactions. This is realized by not allowing the solute and those portions of the membrane in contact with the solute to fit together.

Other significant descriptors of eqs 6 and 7 are the following: ΔE_{TT}(total), the change in total potential energy of the solute−membrane system upon uptake of the solute molecule, $E_{SS}(1-4)_{free}$ and $E_{TT}(1-4)$, the 1−4 nonbonded intramolecular energy of the free solute and the total DMPC−solute complex, respectively, and E_{TT}(vdw), the total van der Waals interaction energy of the membrane−solute complex. All four of these descriptors reflect the molecular flexibility of the solute and/or solute−membrane complex, and have roles similar to those of ΔE_{TT}(hb) and T_G, as described above, in the expression of % Abs. Given the relatively large number of descriptors found in eqs 6 and 7 that reflect molecular flexibility, it would seem that % Abs is very sensitive to the molecular flexibility of both the solute and solute−membrane complex. Table 5 shows the linear cross-correlation matrix of the descriptors found in the two models, and it is clear that the descriptors identified as reflecting molecular flexibility are overall, and somewhat surprisingly, not cross-correlated to one another. This lack of cross-correlation can be attributed to these descriptors capturing molecular flexibility with respect to different structural features of the solute, membrane, and their joint interactions. For example, $E_{SS}(1-4)_{free}$ measures the short-range molecular flexibility of the solute due to interacting groups separated by one torsion angle in the solute. In contrast, ΔE_{TT}(hb) reflects the change in molecular flexibility of the entire solute−membrane complex resulting from the overall change in hydrogen bonding in the complex due to uptake of the solute into the membrane.

Dp is the intramolecular dipole moment of the solute in its lowest energy state, and $F(H_2O)$ and PSA are the aqueous free energy of solvation and the polar surface area of the solute, respectively. Each of these three descriptors reflects that as the polarity of the solute increases, that is Dp and PSA increase, and *F*(H2O) becomes more negative, the corresponding absorption of the solute [drug], % Abs, decreases.

From an inspection of the cross-correlation matrix in Table 5, it is seen that PSA and $F(H_2O)$ have, as expected, a high linear correlation since both are measures of polarity. $F(H_2O)$ has been shown to be an important descriptor in MI-QSAR models for Caco-2 permeability.[15] It is likely that the PSA descriptor captures some solvation characteristics of the molecule and acts as a partial "replacement" for $F(H_2O)$ in

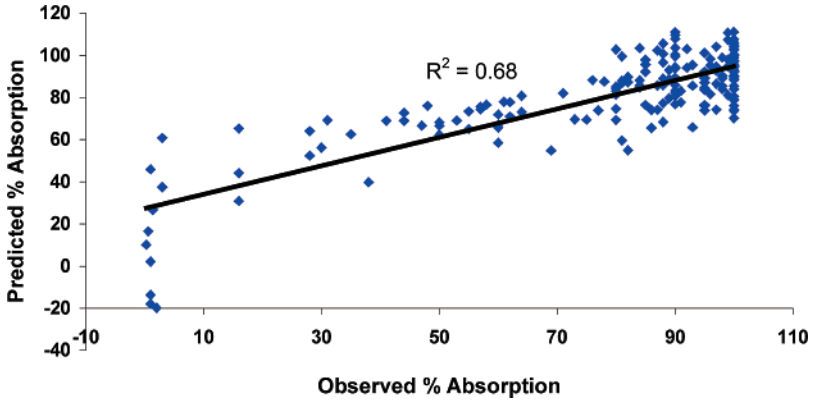

Figure 4. The predicted versus observed percent oral absorption, % Abs, plot for the 188 drugs (training set A) using the MI-QSAR model given by eq 6.

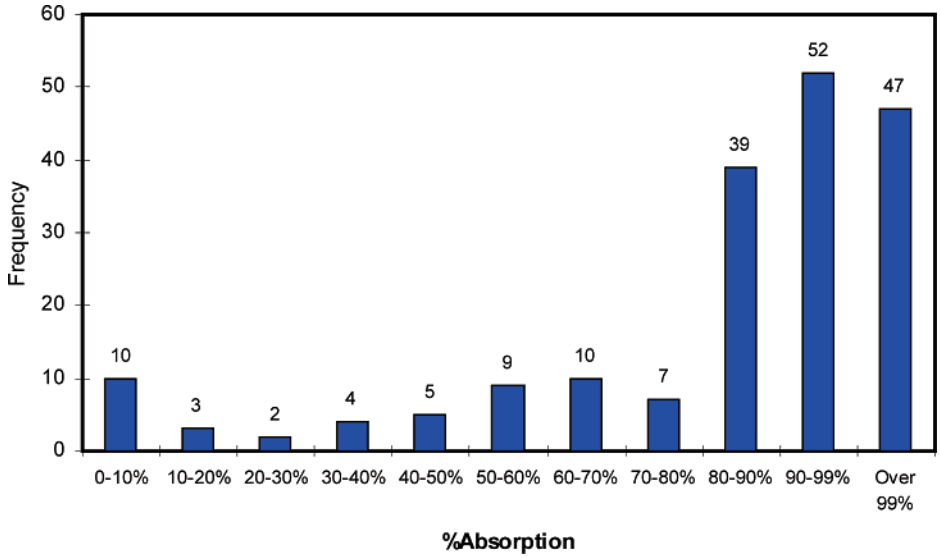

Figure 5. Distribution of % Abs measures across the range of the training set A.

eq 7. The cross-correlation matrix also reveals a high correlation between the 1−4 nonbonded interaction energy within the solute and the same energy term calculated for the entire membrane−solute complex [$E_{SS}(1-4)_{free}$ and E_{TT}-(1−4)], and these descriptors could be playing similar roles in either model.

The 7-term model, eq 6, has a constant term of 78.32 that is very close to the mean percentage absorption (79.77) of the entire training set. This observation is suggestive that eq 6 is a superior statistical model to eq 7, the 6-term model. The predicted versus observed percentage oral absorption, % Abs, plot for eq 6 is shown in Figure 4.

Most molecules in Table 1 [training set A] are drugs that are orally administered. This feature of the data set partially compromises the statistical quality of the data. From a total of 188 compounds, 47 compounds have 100% absorption and 52 compounds are in the range of 90−99% absorption. The molecule with the lowest absorption value is raffinose (0.3%). However, the mean and median absorption values are 80% and 90%, respectively, and the data has a standard deviation of 27%. The distribution of the data in training set A is shown in Figure 5. Given the large size and highly skewed absorption measures across the range of this data set, eq 6 can be judged to be a reasonably significant model even with an $R^2 = 0.68$. Moreover, the absence of any statistical significant models upon random scrambling of the data set also suggests that the model is stable and robust as well.

But, to further investigate the applicability of MI-QSAR descriptors to predict intestinal absorption data, models are also constructed using the kinetic constant (k_{dif}) as the dependent variable. log k_{dif} is calculated using eq 5, but this equation is not defined if FA is 0 or 1, that is, when the absorption percent is either 0% or 100%. Therefore, to derive a log k_{dif} model for training set A, drugs having percentage absorption values of 100% are modified to 99.5% absorption. There are no compounds in the data set with 0% absorption so no corrections on this opposite side of the % Abs range are necessary. The log k_{dif} MI-QSAR model for training set A, with the altered data, is given in Table 6 as eq 8. This model shows slight improvement in quality from the percentage absorption model for the same training set having an R^2 of 0.73 (from 0.68 in eq 6) and a Q^2 of 0.67 (from 0.65 in eq 6).

Table 6. log k_{dif} MI-QSAR Models for Training Sets A, B, and C[a]

eq	*N*	model	R^2	Q^2
8	188	log $k_{dif} = 1.68 + 0.08$ ClogP $- 0.0002$ HOMO $+ 0.006E_{SS}(hb)_{free} + 0.0003T_G - 0.002E_{MS}(vdw+chg) - 0.001E_{TT}(bend)$	0.73	0.67
9	117	log $k_{dif} = -0.12 + 0.006E_{SS}(hb)_{free} + 0.09$ ClogP $- 0.02\ \Delta E_{SS}(bend) - 0.002E_{MS}(vdw+chg)$	0.78	0.74
10	106	log $k_{dif} = -0.44 + 0.10$ ClogP $- 0.002\ \Delta E_{TT}(bend) + 0.0005T_G + 0.005E_{SS}\ (hb)_{free} - 0.001E_{MS}(vdw+chg)$	0.82	0.79

[a] Drugs with 100% and 0% absorption are eliminated in training set B, and zwitterionic drugs are also eliminated in forming training set C. The test set compounds are also eliminated for training sets B and C.

Table 7. Cross-Correlation Matrices of the Descriptors of the MI-QSAR log k_{dif} Model for Training Sets A, B,and C

A. For Training Set A

	ClogP	$E_{SS}(hb)_{free}$	$E_{MS}(vdw+chg)$	HOMO	$E_{TT}(bend)$	T_G
ClogP	1.000					
$E_{SS}(hb)_{free}$	0.493	1.000				
$E_{MS}(vdw+chg)$	0.295	0.362	1.000			
HOMO	0.009	0.002	0.010	1.000		
$E_{TT}(bend)$	0.006	0.013	0.002	0.000	1.000	
T_G	0.005	0.014	0.000	0.008	0.005	1.000

B. For Training Set B

	$E_{MS}(vdw+chg)$	$E_{SS}(hb)_{free}$	$\Delta E_{SS}(bend)$	ClogP
$E_{MS}(vdw+chg)$	1.000			
$E_{SS}(hb)_{free}$	0.353	1.000		
$\Delta E_{SS}(bend)$	0.000	0.004	1.000	
ClogP	0.246	0.589	0.011	1.000

C. For Training Set C

	ClogP	$\Delta E_{TT}(bend)$	$E_{SS}(hb)_{free}$	$E_{MS}(vdw+chg)$	T_G
ClogP	1.000				
$\Delta E_{TT}(bend)$	0.004	1.000			
$E_{SS}(hb)_{free}$	0.610	0.006	1.000		
$E_{MS}(vdw+chg)$	0.247	0.000	0.354	1.000	
T_G	0.147	0.009	0.058	0.020	1.000

Next, in order to probe the effect of the skewed absorption data on the construction of the MI-QSAR models, the 47 drugs with 100% absorption are eliminated from the original data set, that is, training set A. A new data set, divided into a modified training set (B) consisting of 117 compounds and a test set consisting of 24 compounds, is constructed. The test set is selected to span the entire range of the training set, and with the same skewed distribution with respect to oral absorption measures as the training set. In addition, the distribution of relative molecular similarity across the test set is made to be approximately the same as that of the training set. Moreover, in order to determine the effect of the zwitterionic compounds on the QSAR models, the zwitterions are eliminated from the modified training set B to create another training set (C) with 106 drugs. The test set mentioned above has three zwitterionic drugs that are also eliminated to form a distinct test set for training set C. The log k_{dif} MI-QSAR models for training sets B and C are listed in Table 6 along with their respective R^2 and Q^2 values.

The significant descriptors appearing in the resultant MI-QSAR models (eqs 8−10, Table 6) are ClogP, the free space intramolecular solute hydrogen-bonding energy [$E_{SS}(hb)_{free}$], the change in intramolecular bending energies of the solute and the total membrane−solute complex upon uptake of the solute, [ΔE_{SS}(bend) and ΔE_{TT}(bend), respectively], the total bending energy of the membrane−solute complex [E_{TT}(bend)], the sum of intermolecular van der Waals and electrostatic energies between the phospholipid and the "bound" solute molecule [E_{MS}(vdw+chg)], the highest occupied molecular orbital energy [HOMO], and the hypothetical glass transition temperature of the solute [T_G].

ΔE_{SS}(bend), ΔE_{TT}(bend), and [E_{TT}(bend)] are descriptors again reflective of molecular flexibility and play the same role in eqs 8−10 as ΔE_{TT}(total) and similar descriptors do in eqs 6 and 7. E_{MS}(vdw+chg) is a direct estimate of the sum of the electrostatic and hydrogen bonding taking place between the membrane and the solute. log k_{dif} is predicted to modestly increase with increasing membrane−solute electrostatic and hydrogen bonding [more negative values of E_{MS}(vdw+chg)]. This relationship would suggest that solute partitioning into the membrane from solution, and subsequent diffusion, is facilitated by electrostatic and hydrogen bonding between the membrane and the solute.

Table 7 (parts A, B, and C) shows the cross-correlation matrix of the descriptors of eqs 8−10. No significant correlation is present among the descriptors, indicating that each descriptor provides unique information to account for the behavior of the training set data.

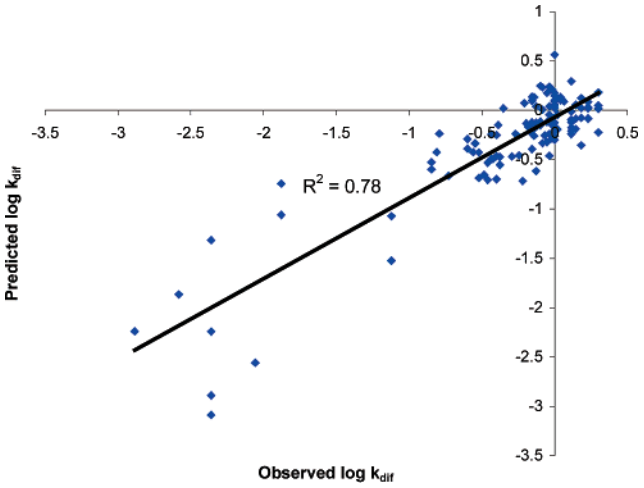

Figure 6. log k_{dif} values for training set B ($N = 117$) observed and as predicted by the MI-QSAR model (eq 9).

Table 8. Observed and Predicted log k_{dif} Values, and Corresponding Residuals of Fit, for the Compounds of the Test Sets

				log k_{dif}			
				predicted		residual	
	molecule	% Abs	obsd	eq 9	eq 10	eq 9	eq 10
1	naproxen	99	0.301	0.111	0.403	−0.190	0.102
2	minoxidilne	98	0.230	−0.326	−0.056	−0.556	−0.287
3	disulfiram	97	0.183	0.201	0.112	0.018	−0.071
4	codeine	95	0.114	−0.035	−0.197	−0.149	−0.311
5	oxprenolol	95	0.114	−0.075	−0.097	−0.189	−0.212
6	amrinone	93	0.063	−0.350	−0.233	−0.412	−0.296
7	amphetamine	90	0.000	−0.031	0.008	−0.031	0.008
8	nisoldipine	90	0.000	0.317	0.329	0.317	0.329
9	dihydrocodeine	89	−0.018	−0.117	−0.171	−0.098	−0.152
10	nitrendipine	88	−0.036	0.076	0.141	0.111	0.177
11	lansoprazole	85	−0.084	0.107	0.130	0.191	0.214
12	captopril	84	−0.099	−0.107	−0.142	−0.008	−0.043
13	flecainide	81	−0.142	−0.075	0.100	0.067	0.242
14	ethambutol	80	−0.156	−0.336	−0.410	−0.181	−0.254
15	famciclovir	77	−0.195	−0.326	−0.303	−0.131	−0.108
16	cimetidine	64	−0.353	−0.279	−0.032	0.074	0.321
17	reproterol	60	−0.400	−0.450	−0.389	−0.050	0.011
18	atenolol	50	−0.521	−0.427	−0.400	0.095	0.121
19	famotidine	38	−0.683	−1.231	−0.668	−0.549	0.014
20	adefovir	16	−1.121	−0.242	−0.392	0.878	0.729
21	ouabain	1.4	−2.213	−1.332	−1.230	0.881	0.983
22	amoxicillin	93	0.063	−0.683	zwitterion	−0.746	
23	levodopa	86	−0.069	−0.685	zwitterion	−0.616	
24	methyldopa	41	−0.640	−0.662	zwitterion	−0.022	

The MI-QSAR model for training set B (eq 9) exhibits a better statistical significance than do the models for the original training set A (eqs 6 and 7). log k_{dif} values for training set B, as predicted by the model expressed as eq 9, are plotted in Figure 6. Additional validation of the model (eq 9) is performed using the test set described above, and given in Table 8, which spans the entire log k_{dif} range of training set B. Equation 9 performs marginally in predicting the log k_{dif} values of the test set. The correlation (R^2) between

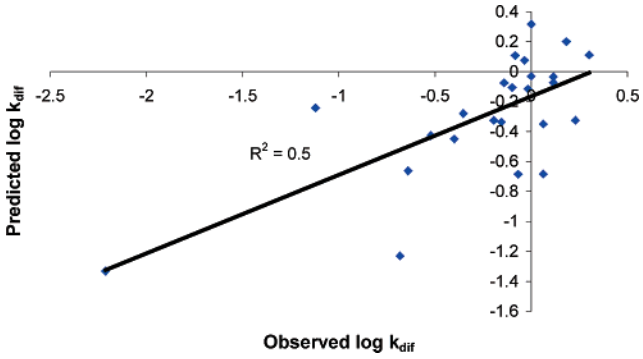

Figure 7. log k_{dif} values for the test set ($N = 24$) observed and as predicted by the MI-QSAR model (eq9).

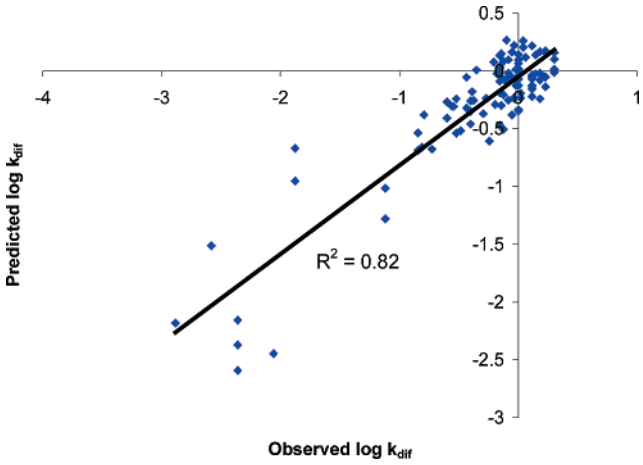

Figure 8. log k_{dif} values for training set C ($N = 106$) observed and as predicted by the MI-QSAR model (eq 10).

predicted and observed log k_{dif} values for the test set is 0.5 (plotted in Figure 7 and tabulated in Table 8). As a diagnostic check to evaluate model predicitivty, eliminating the two largest outliers, compounds 20 and 21 of Table 8, from the test set improves the predictive R^2 value to only 0.6. This suggests that the limitations in the accurate predictivity of eq 9 are distributed reasonably evenly across the test set.

Eliminating the zwitterions from training set B significantly improves the statistical quality of the resultant log k_{dif} MI-QSAR model. When the zwitterions are eliminated to form training set C, the cross-validated correlation coefficient (0.79) of the corresponding MI-QSAR model (eq 10) is higher than that of eq 9 (0.74). Thus, it appears that a substantial source for the lack of fit of eq 8 is the inclusion of the zwitterions in training set B. However, more significant is the finding of the solid performance of the log k_{dif} MI-QSAR model, given by eq 10, in predicting the log k_{dif} of the test set molecules. This model predicts the test set, composed of 21 compounds, with good accuracy ($R^2 = 0.70$), which is not too much less than the fit of the MI-QSAR model to the compounds of training set C. Removing two outliers from the test set further improves the predicted R^2 to 0.74. The predicted versus observed log k_{dif} plots of the training and test sets for eq 10 are shown in Figures 8 and 9, respectively.

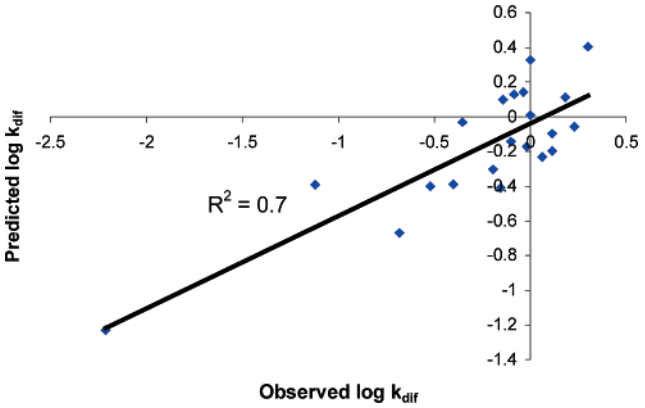

Figure 9. log k_{dif} values for the test set ($N = 21$) observed and as predicted by MI-QSAR model (eq10).

Discussion

In evaluating the QSAR analyses carried out in the work reported here it is important to keep in mind at the outset that gastrointestinal drug absorption is a highly complex process. Thus, it is expected to be quite difficult, if not impossible, to account for all the involved factors in a single QSAR model. However, some important aims and corresponding inferences regarding the nature of drug absorption could be reliably considered as part of this study.

One aim of this study is to ascertain if the MI-QSAR methodology would be applicable to a large, structurally diverse, data set. Most ADME training sets involve many more molecules of higher structural diversity than is found in a typical QSAR training set like enzyme inhibition by a set of analog inhibitors. The ADME data set used in this study is additionally challenging since most of the molecules are drugs that have good oral absorption. Hence, the percent absorbed measures are skewed toward highly absorbed molecules, and nonuniformly influence the data-fitting process of QSAR model building. Overall, the resultant models, as given by eqs 6 − 10, indicate the following:

(1) Only marginal models can be built for the entire data set [eqs 6 and 7] with respect to accuracy, but these models are stable and significant as judged by the R^2 and Q^2 of each model being nearly identical to one another. Thus, the descriptors of these models may meaningfully reflect the mechanism of drug absorption.

(2) Accurate QSAR models could only be built after some data pruning. Elimination of the many (47) compounds reported to have 100% absorption, as well as zwitterionic compounds, led to an accurate and predictive model as expressed by eq 10. One can argue that the many compounds with 100% absorption in the training set unduly bias the fitting of the data and lead to distorted models, while the zwitterionic compounds may act by a modified/different mechanism of transport making the development of a single QSAR model for a single mechanism of transport difficult, or even meaningless to pursue.

(3) Intermolecular MI-QSAR descriptors are found to play a vital role in describing human intestinal oral absorption. A composite examination of the final set of best MI-QSAR models leads to the conclusion that "classic" intramolecular QSAR descriptors are not adequate to describe intestinal absorption. It is emphasized that the identical intramolecular QSAR descriptors found to be significant in other reported absorption and distribution ADME QSAR models were included in the set of trial descriptors of this study. However, several of these intramolecular descriptors were not as important relative to intermolecular MI-QSAR descriptors in building the best models. For example, polar surface area, PSA, is found in many "intramolecular" ADME QSAR models reported in the literature, but only appears once, that being in eq 7, in this work.

(4) Equations 6−10 can be generalized to a form involving three types of thermodynamic processes:

$$\begin{aligned}[\%\ \text{Abs}]\ \text{or}\ [\log k_{dif}] = {} & (\text{a constant value}) + \\ & (\text{solubility and partitioning}) \\ & + (\text{membrane−solute binding}) \\ & + (\text{conformational flexibility of the} \\ & \quad \text{solute and/or membrane}) \quad (11)\end{aligned}$$

Table 9 reports how the descriptors of eqs 6−10 are distributed with respect to these three types of thermodynamic processes. An inspection and comparison of eqs 6 − 10 suggests that % Abs is dependent upon both aqueous−membrane partitioning and aqueous solubility of the drug [eqs 6 and 7], while the associated diffusion process of absorption, as represented by log k_{dif}, is largely governed by aqueous−membrane partitioning as ClogP is only found in eqs 8−10. In making these assessments it is remembered that ClogP is not an explicit measure of aqueous and/or membrane solubilities, but rather an approximate measure of their ratio.

Diffusion is seemingly only influenced by direct membrane−solute "binding", while overall drug absorption, % Abs, involves not only direct membrane−solute interactions but also interactions influencing structural reorganization of the membrane. Finally, there are no apparent differences in the types, or sources, of conformational flexibility of the drug and/or membrane with respect to % Abs and log k_{dif}. The same types of descriptors reflecting molecular flexibility are found in eqs 6 and 7 as in eqs 8−10 as can be seen in Table 9.

Most papers reporting QSAR models for transport ADME properties do not explicitly discuss these models in terms of thermodynamic processes. However, discussions around some of these literature models suggest that constraints on molecular lipophilicity and polar surface area are necessary for effective barrier transport.

A second aim of this study is to determine, as far as possible from the MI-QSAR models, how similar the Caco-2 cell permeation process[30] is to human intestinal oral drug absorption. Caco-2 cell permeation has long been used as a

(30) Pinto, M.; Robine-Leon, S.; Appay, M.; Kedinger, M.; Triadou, N.; et al. Caco-2 cell monolayer a surrogate marker for in vivo intestinal permeability in humans. *Biol. Cell* **1983**, *47*, 323−328.

Table 9. The Distribution of the MI-QSAR Descriptors of Eqs 6−10, 12, and 13 with Respect to Aqueous Solubility, Membrane−Solute Interaction/Binding, and Solute Conformational Flexibility in the Membrane

eq	solubility and partitioning	membrane−solute interactions	solute and membrane conformational flexibility
6	ClogP; $F(H_2O)$	ΔE_{TT}(hb); ΔE_{TT}(total); Dp	$E_{SS}(1-4)_{free}$; T_G
7	ClogP; PSA	ΔE_{TT}(hb); ΔE_{TT}(vdw)	$E_{TT}(1-4)_{free}$; T_G
8	ClogP	HOMO; E_{MS}(vdw+chg)	E_{TT}(bend); T_G; $E_{SS}(hb)_{free}$
9	ClogP	E_{MS}(vdw+chg)	ΔE_{SS}(bend); $E_{SS}(hb)_{free}$
10	ClogP	E_{MS}(vdw+chg)	ΔE_{TT}(bend); $E_{SS}(hb)_{free}$; T_G
12 Caco-2 cell permeation, ref 15	$F(H_2O)$	ΔE_{TT} (hb)	$E_{TT}(1-4)_{free}$; $E_{SS}(hb)_{free}$
13 BBB penetration, ref 16	ClogP; PSA	E_{MS}(chg+hb)	$E_{TT}(1-4)_{free}$; $E_{SS}(tor)_{free}$

laboratory model for oral drug absorption.[31] We previously developed an MI-QSAR Caco-2 cell permeation model[15] which is given by

$$P_{Caco-2} = -14.62 + 0.71F(H_2O) + 0.07\Delta E_{TT}(hb) - 0.26E_{SS}(hb) + 0.06E_{TT}(1-4) \quad (12)$$

$$N = 30 \qquad R^2 = 0.82 \qquad Q^2 = 0.75$$

The descriptor terms in eq 12 have been included in Table 9 to facilitate comparisons to both the % Abs and log k_{dif} MI-QSAR models. The descriptor terms of eq 12 are, overall, largely indistinguishable from those of eqs 6−10. However, eq 6 has the largest number of common descriptors to those of eq 12. Hence, based solely on the descriptors of the MI-QSAR models, and indirectly on the mechanism of transport the descriptors likely reflect, it is reasonable to conclude that intestinal absorption and Caco-2 cell permeation involve similar transport processes. The absence of a ClogP term in eq 12, however, does suggest that water−membrane partitioning may be less important in Caco-2 cell permeation than in human intestinal oral absorption.

The regression coefficients of eqs 6−12 have not been normalized with respect to their weightings within a given MI-QSAR model. Still, the relative values of the regression coefficients of eq 12 can be qualitatively compared to those of eqs 6 and 7. The two training sets are quite similar with respect to both chemical structures of the molecules of the training sets and the corresponding range/magnitude of the dependent variables [% Abs and P_{Caco-2}]. Such a qualitative comparison suggests that increasing aqueous solubility [an increasingly negative $F(H_2O)$ value] of a drug more significantly decreases Caco-2 cell permeation [regression coefficient = 0.71] than intestinal absorption [regression coefficient = 0.31, eq 6]. Minimizing the disruption in the overall hydrogen bonding of both the membrane and drug upon the uptake of the drug into the membrane, as measured by ΔE_{TT}(hb), maximizes both Caco-2 cell permeation and intestinal absorption. However, this factor is again more significant in Caco-2 cell permeation than for intestinal permeation.

A third aim of this study is to compare the descriptor terms of a blood−brain-barrier (BBB) penetration MI-QSAR model to the intestinal absorption MI-QSAR models, eqs 6−10, as well as to the Caco-2 cell permeation model, eq 12. Previously, we developed an MI-QSAR model for BBB penetration[16] that is given by

$$\log BBB = 0.0156 - 0.0231\ PSA + 0.1591\ ClogP - 0.0071E_{MS}(chg + hbd) + 0.0346E_{SS}(tor) + 0.0075\Delta E_{TT}(1-4) \quad (13)$$

$$N = 56 \qquad R^2 = 0.845 \qquad Q^2 = 0.795$$

The descriptors of eq 13 are also listed in Table 9 to readily permit comparisons among the descriptors of the various MI-QSAR models. The descriptors of eq 13 match up reasonably well to those of eqs 6−10, as well as those of eq 12 for Caco-2 cell permeation. Moreover, eq 13 includes a ClogP term which could be argued makes it more similar overall to eqs 6−10 than to eq 12. But while available experimental data indicates that BBB penetration exhibits a trend with oral drug absorption, and also with Caco-2 cell permeation, it is less indicative of human intestinal oral absorption than is Caco-2 cell permeation. Thus, solute differences between BBB penetration and oral absorption are expected to be reflected in the specific descriptors not in common between the BBB and oral absorption models, and/or differences in the relative importance, as measured by the regression coefficients, of the common descriptors in the MI-QSAR models for these two transport processes.

Again, like eq 12, the relative values of the regression coefficients of eq 13 can be qualitatively compared to those of eqs 6−10 for identical descriptors since the training sets are quite similar. However, such comparisons are most reasonable for eqs 8−10, where the common range in the dependent variables is about the same as that for eq 13 [log BBB, log k_{dif}]. Moreover, a comparison between eq 10 and eq 13 is particularly appropriate because the R^2 values of these two MI-QSAR models are also about the same.

One immediate observation in comparing BBB penetration and human intestinal oral absorption models is that the regression coefficient of the ClogP term for the BBB MI-QSAR model is about twice as large as the regression coefficients in eqs 8−10. Thus, BBB penetration is predicted to be more sensitive to increasing drug lipophilicity than

(31) Artursson, P. Cell cultures as models for drug absorption across the intestinal mucosa. *Control Rev. Ther. Drug Syst.* **1991**, *8*, 305−330.

intestinal oral absorption. Moreover, the BBB penetration and intestinal oral absorption models all contain an identical drug−membrane binding descriptor, E_{MS}(chg + hbd). The E_{MS}(chg + hbd) term for BBB penetration has a regression coefficient at least three times larger in magnitude than those of eqs 8−10. As a result, BBB penetration is predicted to be influenced more significantly by binding to the membrane than is intestinal oral absorption.

The remaining descriptors of eq 13 are different from those of eqs 8−10 and, with the exception of PSA, are simply different representations of drug and/or membrane molecular flexibility. Thus, we would conclude that the BBB barrier is very sensitive to the molecular flexibility of both itself and the drug passing through it, in a considerably different way from that of the oral absorption barrier.

Equations 1−5 that define the first-order process of passive diffusion also provide a useful method to estimate diffusion constants. However, it must be pointed out that these equations do not take into consideration the physiology of the gastrointestinal tract. In reality, a drug experiences different pH environments in the stomach and intestine, and the transit time is also variable.[29] The underlying assumption in deriving these models is that the intestinal transit time is constant for all molecules of the data set. Thus, while a diffusion rate constant is estimated for the purpose of this study, it may not be reliable and/or accurate to define the process of absorption based upon such a correspondingly simple kinetic rate equation.

Acknowledgment. This work was funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant 1 R21 GM075775-01. Information on Novel Preclinical Tools for Predictive ADME-Toxicology can be found at http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-04-023.html. Links to nine initiatives are found here: http://nihroadmap. nih.gov/initiatives.asp. This work was also supported, in part, by The Procter & Gamble Company. Resources of the Laboratory of Molecular Modeling and Design at UIC, and at The Chem21 Group, Inc., were used in performing these studies. We are grateful to Tony Hopfinger, Jr. for editing and proofreading this manuscript.

MP0600900