Rapid Calculation of Polar Molecular Surface Area and Its Application to the Prediction of Transport Phenomena. 1. Prediction of Intestinal Absorption

DAVID E. CLARK*

Contribution from *Rhône-Poulenc Rorer, Dagenham Research Centre, Rainham Road South, Dagenham, Essex, RM10 7XS, United Kingdom.*

Received October 7, 1998. Final revised manuscript received April 22, 1999.
Accepted for publication April 30, 1999.

Abstract □ A method for the rapid computation of polar molecular surface area (PSA) is described. It is shown that consideration of only a single conformer when computing PSA gives an excellent correlation with intestinal absorption data—as good as previously reported methods employing multiple conformers. Circumventing a time-consuming conformational analysis opens the possibility of computationally screening large numbers of compounds for problems relating to absorption prior to synthesis. The robustness of the criterion for identifying poorly absorbed compounds (PSA $\geq$ 140 Å^2) is illustrated through its application to a diverse test set of 74 drugs. The PSA-based method is also compared to an experimental method for absorption prediction recently described in the literature.

Introduction

In recent years, the pharmaceutical industry has come under increasing pressure to reduce the time that is required for the discovery and development of novel therapeutics. The advent of combinatorial chemistry and high-throughput screening methodologies has brought with it the ability to synthesize and evaluate orders of magnitude more compounds than has been possible using traditional means. However, these new high-throughput techniques alone do not provide the means to faster drug discovery. Increasingly, researchers have become aware that it is not merely the *number* of compounds made and tested that is important; the *nature* of those compounds is just as vital.

In particular, the focus of drug discovery is now not simply on achieving the best possible potency against the biological target of interest, but also on seeking favorable ADME (absorption, distribution, metabolism, and excretion) properties and performing these two tasks in parallel, rather than in sequence.[1−4] The emphasis is now on "failing fast", i.e., weeding out compounds with poor physicochemical properties early in the drug discovery phase, thereby saving both time and expense.

To this end, a battery of in vitro experimental methods has arisen to help screen candidate molecules for their ADME characteristics.[1,5] Of particular note are Caco-2 monolayers for predicting in vivo intestinal absorption[6,7] and systems for the evaluation of metabolic susceptibility employing human liver microsomes, hepatocytes, or recombinant P450 isozymes.[8] The use of artificial membranes for evaluating absorption processes has also been recently reported.[9] More rapid in vivo tests are also being developed such as "cassette" dosing protocols in which multiple compounds are administered in a single dose to a single animal.[10] However, all these techniques require the synthesis of the compounds to be tested. Even more time and effort could be saved if it were possible for computational techniques to assess reliably compounds for ADME properties prior to synthesis and to identify those likely to be problematic. Such problem compounds could then be rejected or assigned a reduced priority for synthesis. For this reason, several groups have developed computational screening methods seeking to distinguish between druglike and nondruglike compounds in a general sense.[11−13]

For most drugs, the preferred route of administration is by oral ingestion. Researchers have therefore sought to delineate the physicochemical properties that favor intestinal absorption[14,15] and to develop computational methods for its prediction. The so-called "rule-of-5" has proved very popular as a rapid, if approximate, screen for compounds likely to be poorly absorbed.[16] This rule states that if a compound satisfies any *two* of the following rules, it is likely to exhibit poor intestinal absorption:

- Molecular weight > 500
- Number of hydrogen bond donors > 5 (a donor being any O−H or N−H group)
- Number of hydrogen bond acceptors > 10 (an acceptor being any O or N including those in donor groups)
- Calculated log P > 5.0 (if ClogP[17] is used) or > 4.15 (if MlogP[18] is used)

Wessel et al.[19] recently reported the generation of a QSPR (quantitative structure−property relationship) model from a training set of 76 compounds with human fractional absorption (%FA) data using a genetic algorithm with a neural network scoring function. The errors (RMSE) in prediction from this model were 9.4% for the training set and 16.0% for the test set of 10 compounds. Another regression-based approach was described by Hirono et al.;[20] however, their work used oral bioavailability data which may include metabolism effects and so cannot be compared directly to intestinal absorption.

Polar Surface Area

The use of molecular surface areas in the modeling of solvation and partitioning processes has a long history.[21,22] More recently, approaches to absorption prediction have been developed that involve a quantity derived from the molecular surface known as the *polar surface area* (PSA). The PSA of a molecule is defined as the area of its van der Waals surface that arises from oxygen or nitrogen atoms or hydrogen atoms attached to oxygen or nitrogen atoms. As such, it is clearly related to the capacity of a compound to form hydrogen bonds. One study relating PSA to

* Corresponding author. Phone: +44 181 919 3353. Fax: +44 181 919 2029. E-mail: david-e.clark@rp-rorer.co.uk.

10.1021/js9804011 CCC: $18.00
Published on Web 06/22/1999

intestinal absorption was that of van de Waterbeemd et al.,[23] in which a quantitative structure−absorption relationship was derived for the passage of 17 compounds across a Caco-2 monolayer:

$$\log P_{app} = 0.008(\pm 0.002)\mathrm{MW} - 0.043(\pm 0.008)\mathrm{PSA} - 5.165(\pm 0.605) \quad (1)$$

where $\log P_{app}$ is the logarithm of the apparent permeability through the monolayer (in cm/s), and MW is the molecular weight of the compounds. The standard errors of the regression coefficients are given in parentheses. The correlation coefficient for this equation was $r = 0.833$ ($r^2 = 0.694$).

The method used by van de Waterbeemd et al.[23] to calculate PSA considered only a single conformation of the molecules concerned. By contrast, Palm and co-workers employed a measure termed the "dynamic" PSA (PSA_d), which is a Boltzmann-weighted average value computed from an ensemble of low-energy conformers obtained by a detailed conformational search.[24] This kind of Boltzmann averaging of molecular surface areas seems to have been used first by Hermann in the modeling of hydrocarbon solubility[21] and has been applied more recently by Lipkowitz et al.[25] in a study of weakly bound diastereomeric complexes. In their work, Palm et al. showed that PSA_d correlated well with intestinal absorption, in terms of both measurements of Caco-2 monolayer permeability[24] and the fraction absorbed (%FA) in humans.[26] In the former case, the correlation between PSA_d values and Caco-2 absorption was $r^2 = 0.99$. In the latter study on 20 carefully selected drugs, a sigmoidal fit with an $r^2 = 0.94$ (RMSE = 9.2%) was obtained. From this sigmoidal curve, Palm et al.[26] suggested that molecules with a PSA_d of ≥ 140 Å^2 should exhibit a %FA of < 10% and that this PSA_d value could therefore be used to identify poorly absorbed compounds prior to synthesis. Conversely, completely absorbed molecules (%FA > 90%) exhibited PSA_d values of ≤ 60 Å^2.

Palm et al.[24] argued that their PSA_d measure, by taking account of multiple low-energy conformations, should give a better description of molecular surface properties than methods that consider only a single conformer. A similar argument has been advanced recently by Krarup and co-workers[27] who developed a PSA_d measure similar to that described by Palm et al.,[24] but using the solvent-accessible, rather than the van der Waals, molecular surface. The difficulty with these "dynamic" methods from a practical point of view is that they are very computationally expensive. To carry out a conformational search with energy minimization for even a moderately flexible small molecule can take several hours of CPU time on a modern workstation. Clearly, a dynamic PSA calculation becomes impractical if one wants to consider more than just a few molecules on a routine basis. Furthermore, from a theoretical viewpoint, the use of gas-phase conformational energies to calculate the Boltzmann-averaged PSA_d value might be questioned on the grounds that, in solution, solvation effects could significantly alter the relative energies of the conformers. To counter this latter argument, it should be noted that a recent paper by Palm et al.[28] showed that, for most of the set of nine beta-blockers studied, the simulation environment (gas phase, water, or chloroform) had only a small effect on the value of PSA_d.

Despite being advocates of a "dynamic" approach, it is noteworthy that both Palm et al.[24] and Krarup et al.[27] concede that good correlations with absorption can be obtained by just considering a single conformation. To quote from the former: "Surprisingly, the correlations between the surface properties of the global minimum conformations and the permeability were *only slightly poorer* [my italics] than the correlations obtained with the dynamic properties".[24] In the work of the latter group,[27] an excellent correlation ($r^2 = 0.98$) with absorption was discovered using just the (extended) starting conformation for the molecular dynamics simulation. These observations accord with earlier work by Pearlman[22] who demonstrated that single conformation-based surface areas could be as effective as the Boltzmann-averaged areas described by Hermann[21] for the modeling of the free energy of cavity formation. These results encouraged us to investigate in more detail the use of calculations of PSA, from just a single conformer, for the prediction of intestinal absorption. If reliable predictions could be made on this basis, it would open the way for the rapid prescreening of large compound collections or combinatorial libraries to eliminate molecules likely to show poor absorption characteristics. These compound sets might already be in existence, perhaps being considered for purchase from an external source, or they could be *virtual*, i.e., not yet synthesized but represented in a computer-readable form. In either case, a computational procedure for assessing absorption in a reasonable time scale would be of great utility in the drug discovery process.

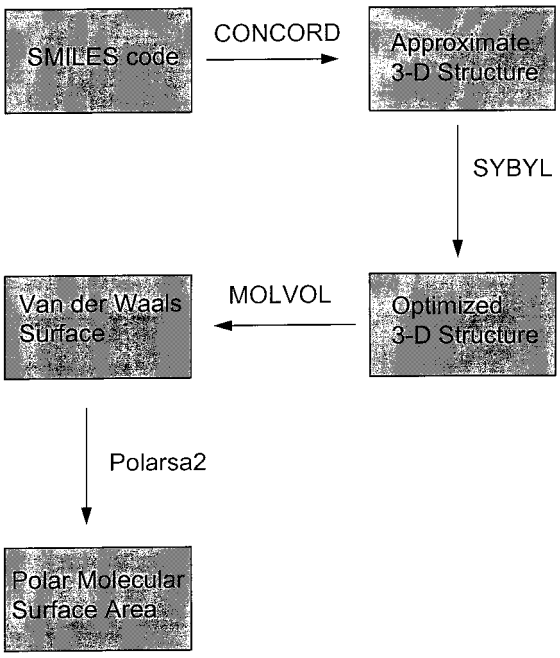

Figure 1—Flowchart of computational procedures.

In the remainder of this paper, the computational methods we have developed will be detailed. Following that, we shall demonstrate their performance first on the training set used by Palm et al.[26] and then on a larger test set drawn from the work of Wessel et al.[19] We shall also compare the results of our method with those of the experimental system developed by Kansy and co-workers.[9] After a discussion of the results obtained to date, some future directions for research in this area will be postulated. The paper immediately following this one will describe the further use of PSA calculations for the prediction of blood-brain partitioning.

Computational Methods

A flowchart of the computational processes employed in this work is given in Figure 1. In detail, the steps are as follows:

1. The molecule is encoded as its neutral species in SMILES[29] format with appropriate stereochemical designations where the stereochemistry is known.

Table 1—Data for the Set of Compounds from Palm et al[26]

compound	PSA/Å^2	PSA$_d$/Å^2	%FA	rule-of-5 status
metoprolol	57.2	53.1	100	pass
nordiazepam	47.5	45.1	99	pass
diazepam	34.5	33.0	97	pass
oxprenolol	53.2	46.8	97	pass
phenazone	28.0	27.1	97	pass
oxazepam	55.6	66.9	97	pass
alprenolol	41.8	37.1	96	pass
practolol	77.2	73.4	95	pass
pindolol	60.9	56.5	92	pass
ciprofloxacin	80.1	78.7	69	pass
metolazone	95.9	94.5	64	pass
tranexamic acid	71.5	69.2	55	pass
atenolol	93.3	90.9	54	pass
sulpiride	101.4	100.2	36	pass
mannitol	129.6	116.6	26	pass
foscarnet	117.3	115.3	17	pass
sulfasalazine	148.6	141.9	12	pass
olsalazine	147.0	141.0	2.3	pass
lactulose	197.8	177.2	0.6	warning
raffinose	266.8	242.1	0.3	warning

2. The program CONCORD[30,31] is used to convert the SMILES representation into an approximate 3-D structure.

3. This conformation is then energy-minimized to relieve any close steric contacts using the maximin2 minimizer in SYBYL.[32] Minimization is terminated after either 1000 iterations or when a gradient of less than 0.05 kcal/(mol.Å) is attained.

4. The minimized conformation is passed to the MOLVOL program developed by Dodd and Theodorou.[33] MOLVOL computes the van der Waals molecular surface area for the conformation and outputs the contributions of the individual atoms to the surface area.

5. Finally, an in-house Fortran program, Polarsa2, sums the contributions of the polar atoms (N, O, and H attached to N or O) and outputs the PSA value.

It should be noted that this procedure is entirely automated by means of a C-shell script and so large numbers of molecules can be processed in batch. An intranet-based interface is also available for medicinal chemists to calculate PSA-based predictions of absorption and blood-brain barrier penetration on single molecules. Typically, the CPU time required to process a compound is of the order of 10−15 CPU seconds (SGI R10000 workstation), most of this time being required for the energy minimization step (which can be curtailed if very large numbers of molecules are to be processed). This efficiency of calculation enables the processing of large sets of compounds, particular virtual combinatorial libraries, in a realistic time frame.

Results

Training Set—The 20 drug molecules studied by Palm et al.[26] were encoded as SMILES and processed in batch using the procedure described above. The CPU time required for this was 138 s (SGI R10000 processor). The resulting PSA values are shown in Table 1, along with the PSA$_d$ and %FA values taken from Palm et al.[26] and the status of the compounds according to the rule-of-5. The fit of the single conformer PSA values to the %FA data using a Boltzmann sigmoidal curve[34] is shown in Figure 2. The statistics for this fit ($r^2 = 0.94$, RMSE = 9.1%) are almost identical to those quoted by Palm et al.[26] ($r^2 = 0.94$, RMSE = 9.2%). Reading from the curve, the PSA value corresponding to a fractional absorption of 10% is 139.4 Å^2 and for 90% absorption, the value is 61 Å^2. Again, these figures are almost identical to those quoted by Palm et al.[26] The correlation coefficient between the values of PSA and PSA$_d$ presented in Table 1 is $r = 0.996$. As might be expected, the agreement in absolute terms is worse for the more-flexible, more-polar compounds such as mannitol, lactulose, and raffinose. Of the more rigid compounds, oxazepam also shows quite a large absolute difference between PSA and PSA$_d$. This difference probably reflects an instance where the use of different computational methods between this work and that of Palm et al.[26] (e.g., for 3-D structure generation and molecular mechanics optimization) has a significant effect on the results.

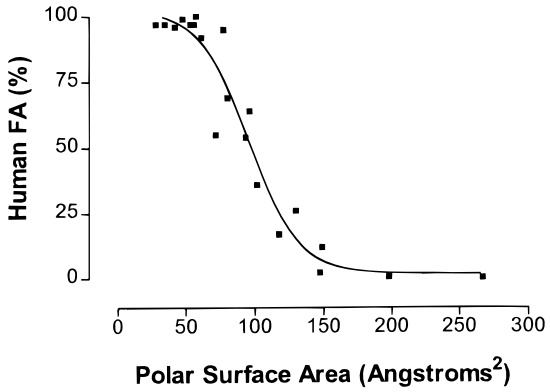

Figure 2—Boltzmann sigmoidal fit of single conformer PSA values to human %FA data for Palm et al.[26] compounds.

From these results, we concluded that, for this set at least, the use of a rapid, single-conformer-based calculation of PSA is sufficient for determining molecules likely to exhibit poor intestinal absorption. It is important to note that all the compounds in the Palm set were chosen on the basis of their being absorbed primarily by passive processes. Thus, the PSA criterion cannot be expected to apply to compounds that are substrates for active transporters.

Applying the value of 140 Å^2 to this set picks out four compounds as being poorly absorbed: sulfasalazine, olsalazine, lactulose, and raffinose. These all show %FA of ≤ 13%. The rule-of-5 only generates warnings for the latter two compounds, suggesting that the PSA value may be a better discriminator of poorly absorbed molecules.

Test Set—To test further the PSA ≥ 140 Å^2 criterion for poor absorption, a further set of compounds was compiled from a recent publication.[19] After removing compounds that are also in the Palm et al. set,[26] 74 compounds remained. PSA values were computed for these compounds and are presented in Table 2 (plotted as a graph in Figure 3) together with the predicted classification of the molecules as "good" (PSA ≤ 61 Å^2), "poor" (PSA ≥ 140 Å^2) or "OK" (140 Å^2 > PSA > 61 Å^2). The rule-of-5 status of each compound is also included, and the %FA values are taken from Wessel et al.[19]

From a practical point of view, when testing this measure, we were most concerned about any *false negative* predictions that might arise from the application of the PSA ≥ 140 Å^2 criterion, i.e., compounds which are predicted to be poorly absorbed but which in fact are well absorbed. False negatives are worrisome because, were the compounds to be discarded on the basis of the prediction, their true absorption would never be discovered. On the other hand, false positive predictions would always be discovered and, while wasteful, would not be serious in the long term. The seven apparent false negatives are highlighted in Table 2 and shown as crosses (×) in Figure 3. Each of these compounds is commented on below.

• Methotrexate has a very high PSA value (225 Å^2), and yet it is still reported as being 100% absorbed. The reason for this is that it is absorbed by a carrier-mediated process which is responsible for folate absorption.[35,36]

• Zidovudine (AZT) just transgresses the PSA limit (142.9 Å^2) while exhibiting 100% absorption. Again, this absorption is made possible by active transport, the carrier in

Table 2—Data for the Set of Compounds from Wessel et al[19]

compound	PSA/Å^2	PSA rating	%FA	rule-of-5 status
acetylsalicylic acid	69.4	OK	100	pass
bumetanide	120.8	OK	100	pass
caffeine	59.2	good	100	pass
corticosterone	75.9	OK	100	pass
desipramine	16.5	good	100	pass
dexamethasone	90.7	OK	100	pass
felodipine	65.2	OK	100	pass
fluvastatin	88.2	OK	100	pass
ibuprofen	42	good	100	pass
ketoprofen	60.9	good	100	pass
loracarbef	117.9	OK	100	pass
lormetazepam	55.9	good	100	pass
methotrexate	**225**	**poor**	**100**	**warning**
ondansetron	38	good	100	pass
prazosin	103.6	OK	100	pass
salicylic acid	62.5	OK	100	pass
testosterone	43.4	good	100	pass
valproic acid	44.2	good	100	pass
zidovudine	**142.9**	**poor**	**100**	**pass**
naproxen	53.8	good	99	pass
prednisolone	98.3	OK	98.8	pass
cephalexin	115.5	OK	98	pass
warfarin	62.8	OK	98	pass
trimethoprim	110.6	OK	97	pass
clonidine	39.5	good	95	pass
fluconazole	75.8	OK	95	pass
imipramine	6	good	95	pass
labetalol	100.6	OK	95	pass
sotalol	90.3	OK	95	pass
amoxicillin	**143.4**	**poor**	**93.5**	**pass**
hydrocortisone	93.7	OK	91	pass
progesterone	41.2	good	91	pass
terazosin	102.3	OK	91	pass
betaxolol	56.3	good	90	pass
chloramphenicol	118.3	OK	90	pass
phenytoin	66.3	OK	90	pass
scopolamine	61.6	OK	90	pass
tenidap	81.9	OK	90	pass
timolol maleate	82.1	OK	90	pass
acebutolol	89.8	OK	89.5	pass
acrivastine	60.2	good	88	pass
trovafloxacin	101	OK	88	pass
bupropion	26.2	good	87	pass
cimetidine	92.1	OK	85	pass
bromazepam	57.4	good	84	pass
methylprednisolone	98.3	OK	82	pass
sorivudine	131.8	OK	82	pass
acetaminophen	58.5	good	80	pass
quinidine	48.6	good	80	pass
cefatrizine	**188.4**	**poor**	**76**	**warning**
guanabenz	72.9	OK	75	pass
propylthiouracil	48.9	good	75	pass
sumatriptan	75	OK	75	pass
lamotrigine	96.4	OK	70	pass
captopril	61.8	OK	67	pass
hydrochlorothiazide	135.2	OK	67	pass
furosemide	127.7	OK	61	pass
ziprasidone	60.5	good	60	pass
etoposide	**183.3**	**poor**	**50**	**warning**
gabapentin	63.4	OK	50	pass
ranitidine	86.7	OK	50	pass
phenoxymethyl penicillinic acid	107.6	OK	45	pass
cefuroxime axetil	**188.6**	**poor**	**36**	**warning**
norfloxacin	81.1	OK	35	pass
nadolol	85.5	OK	34.5	pass
pravastatin	123.9	OK	34	pass
lisinopril	**142.4**	**poor**	**25**	**pass**
chlorothiazide	134.3	OK	13	pass
enalaprilat	115.1	OK	10	pass
cefuroxime	176.8	poor	5	pass
doxorubicin	199.5	poor	5	warning
ganciclovir	151.8	poor	3.8	pass
cromolyn	184.6	poor	0.5	pass
gentamicin	174.4	poor	0	warning

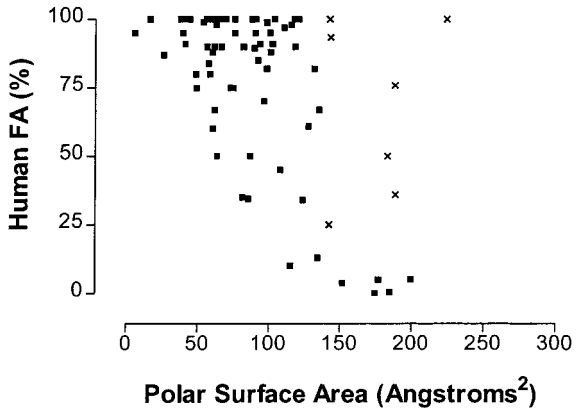

Figure 3—Plot of single conformer PSA values to %FA data for 74 compounds from Wessel et al.[19] set. Seven compounds apparently violating the PSA ≥ 140 Å^2 criterion for poor absorption are shown as crosses (×).

question being a recombinant nucleoside transporter which is responsible for the absorption of pyrimidine nucleosides.[37]

• Amoxicillin and cefatrizine are both amino-β-lactam antibiotics that show good intestinal absorption despite high PSA values. The explanation for this behavior is that both are absorbed via dipeptide carriers.[38,39]

• Cefuroxime axetil shows moderate absorption (36%)—more than would be predicted from its PSA value of 188.6 Å^2. It appears that absorption of this compound takes place by a specialized transport mechanism that obeys Michaelis−Menten kinetics.[40]

• Like zidovudine, lisinopril is a little over the PSA limit and shows a greater than expected absorption. In common with other dipeptide ACE inhibitors, lisinopril is absorbed in a nonpassive process via the dipeptide carrier system.[41]

Thus, all the above compounds are actively transported by one means or another and so are not true false negatives within the scope of the PSA predictions, which only apply to passively transported molecules. The remaining compound, etoposide, is more problematic. To our knowledge, there is no evidence in the literature for any active processes being involved in its intestinal absorption although a recent paper[42] suggests that etoposide distribution into the brain is partly controlled by an active transport process. In any case, etoposide does show very erratic oral bioavailability which has been attributed to low aqueous solubility, slow intrinsic dissolution rate, and chemical instability at pH 1.3[43] and attempts to improve its bioavailability have been unsuccessful.[44] Etoposide, therefore, while apparently a false negative, is not an ideal representative of small molecule oral drugs.

If the above compounds are ignored, the distribution of the remaining points in Figure 3 is roughly sigmoidal; however, there is considerably more scatter than in Figure 2. The reason for this is most likely that the compounds in the Wessel et al. set[19] were not chosen so carefully as those in the Palm et al. set,[26] particularly with regard to the mode of absorption. Even after eliminating the seven "problem" compounds, it is likely that the set contains compounds that are absorbed at least partially by active processes (e.g., acyclovir, cephalexin), and these will tend to add noise to the data.

Nevertheless, the (non)absorption of the remaining 67 compounds is successfully predicted when the PSA value of 140 Å^2 is used to partition the set. These results bolster our confidence that, when passive absorption processes are being considered, the criterion of PSA ≥ 140 Å^2 is a reliable predictor of poorly absorbed compounds and certainly more reliable than the rule-of-5, which only identifies two out of the five compounds showing <10% absorption in this set.

Table 3—Data for the Set of Compounds from Kansy et al[9]

compound	%FA	absorption classification			PSA/Å^2	rule-of-5 status
		expt	PAMPA	PSA		
acetylsalicylic acid	100	h	m	**h**	73.4	pass
alprenolol	93	h	**h**	**h**	41.8	pass
atenolol	54	m	**m**	**m**	93.3	pass
ceftriaxone	1	l	**l**	**l**	220.4	warning
chloramphenicol	90	h	**h**	l	118.4	pass
corticosterone	100	h	**h**	**h**	75.9	pass
coumarin	100	h	**h**	**h**	33.9	pass
dexamethasone	100	h	**h**	m	87.8	pass
diltiazem	92	h	**h**	**h**	60.0	pass
guanabenz	75	h	**h**	**h**	72.9	pass
hydrocortisone	89	h	**h**	m	93.7	pass
imipramine	99	h	m	**h**	6.0	pass
metoprolol	95	h	**h**	**h**	57.2	pass
olsalazine	2	l	**l**	**l**	147.0	pass
propranolol	90	h	**h**	**h**	42.7	pass
salicylic acid	100	h	m	**h**	62.5	pass
sulfasalazine	13	l	**l**	**l**	148.6	pass
sulpiride	35	m	**m**	**m**	102.2	pass
terbutaline	73	h	m	**h**	82.8	pass
testosterone	98	h	**h**	**h**	43.4	pass
theophylline	98	h	m	**h**	74.7	pass
tiacrilast	99	h	m	**h**	79.4	pass
verapamil	95	h	**h**	**h**	73.5	pass
warfarin	93	h	**h**	**h**	62.9	pass

Comparison with a High-Throughput Experimental Method—Recently, Kansy et al.[9] reported the application of a system (called PAMPA − Parallel Artificial Membrane Permeation Assay) for the prediction of passive absorption. PAMPA estimates absorption by measuring the "flux" of a compound through an artificial membrane and is capable of processing large numbers of compounds in parallel by means of 96-well plates. Using the PAMPA system, Kansy et al. classified each member of a set of 25 compounds as being either:

- Highly absorbed (h): %FA ≥ 70, PAMPA flux > 25%
- Moderately absorbed (m): <30 %FA < 70, PAMPA flux 5−25%
- Poorly absorbed (l): %FA ≤ 30, PAMPA flux < 5%

From the curve in Figure 2, the PSA values corresponding to %FA values of 70 and 30 can be determined to be 83 Å^2 and 112 Å^2, respectively. Using these limits, it was decided to compare the classification of the compounds based upon PSA values to that arising from the experimental PAMPA system. Of the 25 compounds comprising the test set for the PAMPA, one (cephalexin) was excluded as it is known to be actively transported (for this reason it was poorly predicted by the PAMPA system). PSA values and rule-of-5 status were calculated for the remaining 24 compounds, and the compounds were classified according to their PSA values. The results are shown in Table 3 where the classifications of the compounds from experiment (human %FA), the PAMPA system, and PSA are collated.

The instances where the PAMPA or PSA predictions are correct are highlighted. From this, it can be seen that the PAMPA makes six classification errors, particularly underestimating compounds likely to be transported by a paracellular route, e.g., theophylline and salicylic acid.[9] The PSA classification system makes only three errors, although one of these (chloramphenicol) is a serious one: the compound being predicted to be in the low absorption class, but in fact having a high absorption (i.e., a false negative). The two other errors in the PSA classification are both steroids (dexamethasone and hydrocortisone) which are predicted to have moderate absorption but which are actually well-absorbed. If the original criterion for poor absorption of 140 Å^2 is applied, all three poor compounds are correctly identified (although two of these are present in the original training set of Palm et al.[26]) and no false negatives are predicted.

Discussion

The results presented above suggest that using single conformer-based calculations to identify compounds with a PSA of ≥ 140 Å^2 provides a rapid and reliable means of eliminating candidates that are likely to be poorly absorbed in the human intestine. This criterion would seem to be more discriminating than the popular rule-of-5[16] and of similar predictive ability to a high-throughput experimental system.[9] The speed of the method makes it suitable for screening large compound sets, and the fact that it does not require the synthesis of the compounds being assessed means it is especially applicable in the design of combinatorial libraries. For example, we are already using PSA values as an additional constraint in our programs for product-based reagent selection.[45]

Our method is rather different in philosophy from the QSPR-type approach adopted by Wessel et al.,[19] inasmuch as we seek simply a rough classification of compounds whereas their model seeks to predict accurate values across the whole range of absorption. While the results of Wessel et al. seem quite reasonable (RMSEs in predicted %FA: training set 9.4%, cross-validation set 19.7%, and external prediction set 16.0%), there are a number of questions about their work, in particular concerning the data set employed. First, as the authors concede, the data set is heavily skewed toward well-absorbed compounds—only 15 of the 86 compounds have %FA values of less than 50. This bias tends to cause their model to be less accurate in predicting %FA values for less-well-absorbed compounds. For instance, enalaprilat is predicted to have %FA of 47.68, whereas its real value is 10%; conversely, lisinopril is predicted to have a %FA of 0, whereas its true value is 25%. Second, and more fundamentally, we would question the constitution of the data set used by Wessel et al. because it contains compounds that are absorbed by active transport processes as well as passively absorbed compounds. To our minds, it seems unrealistic to expect to find a single model that will accurately predict such different physical processes: passive absorption being diffusion controlled while each active transport mechanism requires more specific molecular recognition.[46] It is for this reason that other workers have striven to use only passively absorbed compounds in their studies.[26] Finally, some of the descriptors used in the model require the use of semiempirical molecular orbital (MO) calculations to generate partial charge information, and these can be moderately time-consuming. A more high-throughput model not including the MO-based descriptors was developed for processing combinatorial libraries, but this was found to give less accurate predictions.[19] Nonetheless, it is interesting to examine the six descriptors used in the model described by Wessel et al. Three seem to code for size, flexibility, and shape while the remainder could be related to PSA: SAAA-2 (surface area of hydrogen bond acceptor atoms), SCAA-2 (surface area × charge of hydrogen bond acceptor atoms), and CHDH-1 (charge on donatable hydrogen atoms). This would seem to provide additional evidence of the validity of polar surface measures in the prediction of absorption.

This begs the question: why should PSA be a useful predictor of passive absorption? Two of the key physicochemical determinants of passive absorption are lipophilicity and hydrogen bonding potential,[47] so it would seem

Table 4—Atomic Radii Used in Molecular Surface Calculations

atom	radius/Å
C	1.9
O	1.74
N	1.82
H	1.5
H attached to O	1.1
H attached to N	1.125
S	2.11
P	2.05
F	1.65
Cl	2.03
Br	2.18
I	2.32

likely that PSA describes one of these in some way. However, PSA is not simply related to lipophilicity. This can be clearly shown by considering, for instance, a homologous series of monoalcohols: MeOH, EtOH, PrOH, etc. In this series, the PSA is constant, while the lipophilicity increases with increasing carbon chain length. More generally, for the 86 compounds in the Wessel et al. set,[19] there is only a poor correlation between PSA and ClogP ($r = 0.56$). PSA is more closely related to hydrogen bonding potential as shown by Palm et al.[25] who found a correlation of $r = 0.92$ between PSA_d and a count of the total number of hydrogen bonds capable of being formed by a molecule (H_t). However, PSA is a more subtle descriptor of hydrogen bonding potential than H_t because it can account for 3D effects such as shielding or burial of polar groups by other parts of a molecule, perhaps because of internal hydrogen bonding. Given that calculations of lipophilicity, at least for neutral species, are readily available,[48] combining these with PSA predictions should help to guide drug design toward orally available compounds.

There are a number of factors affecting the PSA value for any given compound. First, it is incontrovertible that PSA will vary with conformation as shown by Palm et al.[24] However, for the large range of compounds studied here, it seems that using a single low-energy conformer as a representative of the ensemble of conformers likely to be present in vivo is a reasonable approximation. While this approximation will tend to hold best for more rigid compounds, more flexible compounds also pose problems to a "dynamic" approach in terms of the increased CPU time required for a thorough conformational search. Second, for compounds with more than one stereocenter, the choice of absolute configuration can also affect the PSA value although, in general, unless the stereochemistry is complex, this is only a minor effect. Finally, the atomic radii and algorithm used for the molecular surface calculation will also have an effect, as will the method of generating the 3-D structure and the particular force field used for any geometry optimization. In this work, we have used the same radii as Palm et al.,[26] with the exception of not distinguishing between sp^2 and sp^3 carbon atoms. The radii used in this work are given in Table 4.[49]

While the PSA criterion for poor absorption seems robust in avoiding false negative predictions, it is not immune from making false positive assessments. A good example of this is pyridostigmine (**1**, Figure 4), which has a very low PSA value but is poorly absorbed. This indicates that a PSA of <140 Å^2 is a necessary, but not sufficient, criterion for absorption. In the case of pyridostigmine, the cause of the poor absorption is the presence of the positively charged quaternary nitrogen which inhibits partitioning into the intestinal membranes. Likewise, the pK_a of a compound is also a key determinant of its absorption; a point well

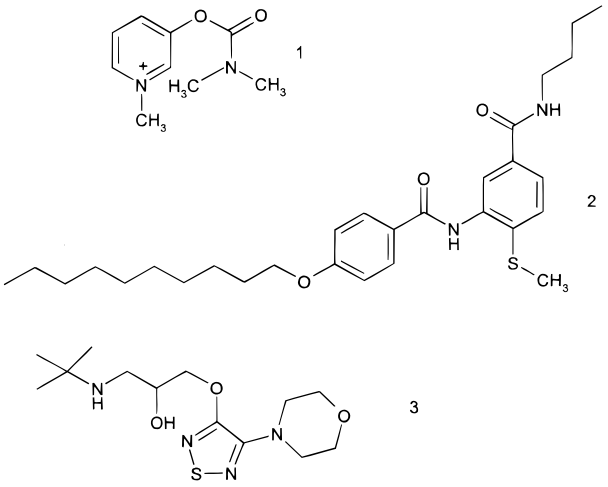

Figure 4—Pyridostigmine (**1**), RP64477 (**2**), and timolol (**3**).

illustrated by a recent publication on 5-HT_{1D} agonists.[50] As indicated above, the lipophilicity of a compound must also be considered. For instance, RP64477 (**2**, Figure 4) has a PSA value of 67.9 Å^2 and on this basis would be expected to be well-absorbed. The fact that it is, in fact, poorly orally available has been attributed to its high log *D* (5.2 at pH 7.4).[51] Finally, aqueous solubility also affects absorption and should be an additional consideration early in compound design.[16] In summary, it is important not to use PSA naively as a guide to absorption but to couple it to existing medicinal chemistry knowledge and intuition. Future work will seek to incorporate some of these factors in an attempt to reduce the number of false positive predictions while avoiding false negatives. Other workers are also exploring these avenues: in a recent article, Winiwarter and coworkers showed that, in addition to PSA, a count of hydrogen bond number and log *D* (octanol/water, pH 5.5 or pH 6.5) were helpful in accurately predicting drug permeability across the human jejunum.[52] Stenberg et al.[53] have also shown recently in a study of a set of peptides that consideration of the nonpolar, as well as the polar, surface area can lead to improved predictions of permeability.

Another interesting avenue of future research is the investigation of refinements to the definition of what constitutes the polar surface area of a molecule. For instance, Krarup et al.[27] reported that for timolol (**3**, Figure 4) and its ester prodrug, a correction was needed to the definition of the polar surface in order to obtain the excellent reported correlation with Caco-2 permeability. The correction was to omit the two nitrogens in the thiadiazole ring from the definition of the polar surface, on the grounds that they are only weak acceptors. When this was done, the correlation between the predicted and actual log P_{app} values improved significantly. The idea of refining the polar surface area so that it takes account of the strength of the hydrogen bonds formed by the groups that comprise it is intriguing. Certainly, the reported improvement in the predictions for timolol and its ester is striking. However, if Krarup et al. were to be consistent, they should also have omitted the non-carbonyl ester oxygen from the polar surface of the timolol prodrug as such oxygens are also known to be extremely infrequent hydrogen-bond acceptors from crystal structure surveys and ab initio calculations.[54–56] The effect of this modification upon their correlation has not yet been investigated. Nonetheless, this idea is certainly worthy of further investigation to see if it helps to improve the accuracy of PSA-based predictions of absorption.

Conclusion

A method for the rapid computation of polar surface area has been described. It has been shown that excellent correlation between polar surface area and intestinal absorption can be obtained without the time-consuming procedure of examining multiple conformations.[57] The use of such rapid calculations in conjunction with the criterion for poor absorption of PSA ≥ 140 Å^2 appears to be an efficient and robust method of computationally screening large numbers of compounds prior to synthesis. It should thus be a valuable tool in both lead generation and optimization.

References and Notes

1. Lloyd, A. W. Can We Relieve the Drug Development Bottleneck? *Pharm. Sci. Technol. Today* **1998**, *1*, 45−46.
2. Rodrigues, A. D. Preclinical Drug Metabolism in the Age of High-Throughput Screening: An Industrial Perspective. *Pharm. Res.* **1997**, *14*, 1504−1510.
3. Lin, J. H.; Lu, A. Y. H. Role of Pharmacokinetics and Metabolism in Drug Discovery and Development. *Pharmacol. Rev.* **1997**, *49*, 403−449.
4. Smith, D. A.; Jones, B. C.; Walker, D. K. Design of Drugs Involving the Concepts and Theories of Drug Metabolism and Pharmacokinetics. *Med. Res. Rev.* **1996**, *16*, 243−266.
5. Tarbit, M. H.; Berman, J. High-throughput Approaches for Evaluating Absorption, Distribution, Metabolism and Excretion Properties of Lead Compounds. *Curr. Opin. Chem. Biol.* **1998**, *2*, 411−416.
6. Artursson, P.; Palm, K.; Luthman, K. Caco-2 Monolayers in Experimental and Theoretical Predictions of Drug Transport. *Adv. Drug Deliv. Rev.* **1996**, *22*, 67−84.
7. Artursson, P.; Borchardt, R. Intestinal Drug Absorption and Metabolism in Cell Cultures: Caco-2 and Beyond. *Pharm. Res.* **1997**, *14*, 1655−1658.
8. Ito, K.; Iwatsubo, T.; Kanamitsu, S.; Nakajima, Y.; Sugiyama, Y. Quantitative Prediction of In Vivo Drug Clearance and Drug Interactions from In Vitro Data on Metabolism, together with Binding and Transport. *Annu. Rev. Pharmacol. Toxicol.* **1998**, *38*, 461−499.
9. Kansy, M.; Senner, F.; Gubernator, K. Physicochemical High Throughput Screening: Parallel Artificial Membrane Permeation Assay in the Description of Passive Absorption Processes. *J. Med. Chem.* **1998**, *41*, 1007−1010.
10. Berman, J.; Halm, K.; Adkison, K.; Shaffer, J. Simultaneous Pharmacokinetic Screening of a Mixture of Compounds in the Dog using API LC/MS/MS Analysis for Increased Throughput. *J. Med. Chem.* **1997**, *40*, 827−829.
11. Gillet, V. J.; Willett, P.; Bradshaw, J. Identification of Biological Activity Profiles Using Substructural Analysis and Genetic Algorithms. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 165−179.
12. Ajay; Walters, W. P.; Murcko, M. A. Can We Learn to Distinguish Between Drug-like and Non-Drug-Like Molecules? *J. Med. Chem.* **1998**, *41*, 3314−3324.
13. Sadowski, J.; Kubinyi, H. A Scoring Scheme for Discriminating between Drugs and Nondrugs. *J. Med. Chem.* **1998**, *41*, 3325−3329.
14. Navia, M. A.; Chaturvedi, P. R. Design Principles for Orally Bioavailable Drugs. *Drug Discov. Today* **1996**, *1*, 179−189.
15. Chan, O. H.; Stewart, B. H. Physicochemical and Drug-Delivery Considerations for Oral Drug Bioavailability. *Drug Discov. Today* **1996**, *1*, 461−473.
16. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3−25.
17. ClogP. Daylight Chemical Information Software, version 4.51. Daylight Chemical Information Inc.: 27401 Los Altos, Suite #370, Mission Viejo, CA 92691.
18. Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull.* **1992**, *40*, 127−130.
19. Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. Prediction of Human Intestinal Absorption of Drug Compounds From Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 726−735.
20. Hirono, S.; Nakagome, I.; Hirano, H.; Matsushita, Y.; Yoshii, F.; Moriguchi, I. Noncongeneric Structure-Pharmacokinetic Property Correlation Studies Using Fuzzy Adaptive Least-Squares: Oral Bioavailability. *Biol. Pharm. Bull.* **1994**, *17*, 306−309.
21. Hermann, R. B. Theory of Hydrophobic Bonding. II. Correlation of Hydrocarbon Solubility in Water with Solvent Cavity Surface Area. *J. Phys. Chem.* **1972**, *76*, 2754−2759.
22. Pearlman, R. S. Molecular Surface Areas and Volumes and their Use in Structure−Activity Relationships. In *Physical Chemical Properties of Drugs*; Medicinal Research Series Vol. 10; Yalkowsky, S. H., Sinkula, A. A., Valvani, S. C., Eds.; Marcel Dekker: New York, 1980; pp 321−347.
23. van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Raevsky, O. A. Estimation of Caco-2 Cell Permeability using Calculated Molecular Descriptors. *Quant. Struct.-Act. Relat.* **1996**, *15*, 480−490.
24. Palm, K.; Luthman, K.; Ungell, A.-L.; Strandlund, G.; Artursson, P. Correlation of Drug Absorption with Molecular Surface Properties. *J. Pharm. Sci.* **1996**, *85*, 32−39.
25. Lipkowitz, K. B.; Baker, B.; Larter, R. Dynamic Molecular Surface Areas. *J. Am. Chem. Soc.* **1989**, *111*, 7750−7753.
26. Palm, K.; Stenberg, P.; Luthman, K.; Artursson, P. Polar Molecular Surface Properties Predict the Intestinal Absorption of Drugs in Humans. *Pharm. Res.* **1997**, *14*, 568−571.
27. Krarup, L. H.; Christensen, I. T.; Hovgaard, L.; Frokjaer, S. Predicting Drug Absorption from Molecular Surface Properties Based on Molecular Dynamics Simulations. *Pharm. Res.* **1998**, *15*, 972−978.
28. Palm, K.; Luthman, K.; Ungell, A.-L.; Strandlund, G.; Beigi, F.; Lundahl, P.; Artursson, P. Evaluation of Dynamic Polar Molecular Surface Area as Predictor of Drug Absorption: Comparison with Other Computational and Experimental Predictors. *J. Med. Chem.* **1998**, *41*, 5382−5392.
29. Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
30. Pearlman, R. S. Rapid Generation of High Quality Approximate 3D Molecular Structures. *Chem. Des. Auto. News* **1987**, *2*, 1−7.
31. CONCORD v4.02. Balducci, R.; McGarity, C. M.; Rusinko, A., III; Skell, J.; Smith, K.; Pearlman, R. S (University of Texas at Austin). Distributed by Tripos, Inc., 1699 S. Hanley Rd., Suite 303, St. Louis, MO 63144-2913.
32. SYBYL 6.4.2. Tripos, Inc., 1699 S. Hanley Rd., Suite 303, St. Louis, MO 63144-2913.
33. Dodd, L. R.; Theodorou, D. N. Analytical Treatment of the Volume and Surface Area of Molecules Formed by an Arbitrary Collection of Unequal Spheres Intersected by Planes. *Mol. Phys.* **1991**, *72*, 1313−1345.
34. A Boltzmann sigmoidal curve follows the equation: $y = \text{bottom} + (\text{top} - \text{bottom})/(1 + \exp((x_{50} - x)/\text{slope}))$. In this instance, y = %FA, x = PSA, and x_{50} is the PSA value at which %FA = 50. Bottom and top refer to the minimum and maximum values of %FA, respectively.
35. Chungi, V. S.; Bourne, D. W. A.; Dittert, L. W. Competitive Inhibition between Folic Acid and Methotrexate for Transport Carrier in the Rat Small Intestine. *J. Pharm. Sci.* **1979**, *68*, 1552−1553.
36. Dudeja, P. K.; Torania, S. A.; Said, H. M. Evidence for the Existence of Carrier-Mediated Folate Uptake Mechanism in Human Colonic Luminal Membranes. *Am. J. Physiol.* **1997**, *272*, G1408−G1415.
37. Yao, S. Y.; Cass, C. E.; Young, J. D. Transport of the Antiviral Nucleoside Analogues 3′-azido-3′-deoxythymine and 2′,3′-dideoxycytidine by a Recombinant Nucleoside Transporter (rCNT) Expressed in Xenopus Laevis Oocytes. *Mol. Pharmacol.* **1996**, *50*, 388−393.
38. Westphal, J. F.; Deslandes, A.; Brogard, J. M.; Carbon, C. Reappraisal of Amoxycillin Absorption Kinetics. *J. Antimicrob. Chemother.* **1991**, *27*, 647−654.
39. Reigner, B. G.; Couet, W.; Guedes, J. P.; Fourtillan, J. B.; Tozer, T. N. Saturable Rate of Cefatrizine Absorption after Oral Administration in Humans. *J. Pharmacokinet. Biopharm.* **1990**, *18*, 17−34.
40. Ruiz-Balaguer, N.; Nacher, A.; Casabo, V. G.; Merino, M. Nonlinear Intestinal Absorption Kinetics of Cefuroxime Axetil in Rats. *Antimicrob. Agents Chemother.* **1997**, *41*, 445−448.
41. Friedman, D. I.; Amidon, G. L. Intestinal Absorption Mechanism of Dipeptide Angiotensin Converting Enzyme Inhibitors of the Lysyl-Prolyl Type: Lisinopril and SQ 29, 852. *J. Pharm. Sci.* **1989**, *78*, 995−998.
42. Burgio, D. E.; Gosl, M. P.; McNamara, P. J. Effects of P-glycoprotein Modulators on Etoposide Elimination and Central Nervous System Distribution. *J. Pharmacol. Exp. Ther.* **1998**, *287*, 911−917.

43. Shah, J. C.; Chen, J. R.; Chow, D. Preformulation Study of Etoposide: Identification of Physicochemical Characteristics Responsible for the Low and Erratic Oral Bioavailability of Etoposide. *Pharm. Res.* **1989**, *6*, 408−412.
44. Joel, S. P.; Clark, P. I.; Heap, L.; Webster, L.; Robbins, S.; Craft, H.; Slevin, M. L. Pharmacological Attempts to Improve the Bioavailability of Oral Etoposide. *Cancer Chemother. Pharmacol.* **1995**, *37*, 125−133.
45. Good, A. C.; Lewis, R. A. New Methodology for Profiling Combinatorial Libraries and Screening Sets: Cleaning Up the Design Process with HARPick. *J. Med. Chem.* **1997**, *40*, 3926−3936.
46. Doring, F.; Will, J.; Amasheh, S.; Clauss, W.; Ahlbrecht, H.; Daniel, H. Minimal Molecular Determinants of Substrates for Recognition by the Intestinal Peptide Transporter. *J. Biol. Chem.* **1998**, *273*, 23211−23218.
47. Conradi, R. A.; Burton, P. S.; Borchardt, R. T. Physicochemical and Biological Factors that Influence a Drug's Cellular Permeability by Passive Diffusion. In *Lipophilicity in Drug Action and Toxicology*; Pliška, V., Testa, B., van de Waterbeemd, H., Eds.; VCH: Weinheim, 1996; pp 233−252.
48. Carrupt, P.-A.; Testa, B.; Gaillard, P. Computational Approaches to Lipophilicity: Methods and Applications. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1997; Vol. 11, pp 241−315.
49. Luthman, K. Private communication, 1998.
50. Castro, J. L.; Collins, I.; Russell, M. G. N.; Watt, A. P.; Sohal, B.; Rathbone, D.; Beer, M. S.; Stanton, J. A. Enhancement of Oral Absorption in Selective 5-HT_{1D} Receptor Agonists: Fluorinated 3-[3-(Piperidin-1-yl)propyl]indoles. *J. Med. Chem.* **1998**, *41*, 2667−2670.
51. Wils, P.; Warnery, A.; Phung-Ba, V.; Legrain, S.; Scherman, D. High Lipophilicity Decreases Drug Transport Across Intestinal Epithelial Cells. *J. Pharmacol. Exp. Ther.* **1994**, *269*, 654−658.
52. Winiwarter, S.; Bonham, N. M.; Ax, F.; Hallberg, A.; Lennernäs, H.; Karlén, A. Correlation of Human Jejunal Permeability (in Vivo) of Drugs with Experimentally and Theoretically Derived Parameters. A Multivariate Data Analysis Approach. *J. Med. Chem.* **1998**, *41*, 4939−4949.
53. Stenberg, P.; Luthman, K.; Artursson, P. Prediction of Membrane Permeability to Peptides from Calculated Dynamic Molecular Surface Properties. *Pharm. Res.* **1999**, *16*, 205−212.
54. Böhm, H.-J.; Brode, S.; Hesse, U.; Klebe, G. Oxygen and Nitrogen in Competitive Situations: Which is the Hydrogen-Bond Acceptor? *Chem. Eur. J.* **1996**, *2*, 1509−1513.
55. Bruno, I. J.; Cole, J. C.; Lommerse, J. P. M.; Rowland, R. S.; Taylor, R.; Verdonk, M. L. Isostar: A Library of Information about Nonbonded Interactions. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 525−537.
56. Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. Ab Initio Study of Hydrogen-Bonded Complexes of Small Organic Molecules with Water. *J. Phys. Chem. A* **1998**, *102*, 3782−3797.
57. During the revision of this paper, van de Waterbeemd et al. published a paper describing the use of single conformation PSA values in the classification of drug compounds as CNS active or CNS inactive: van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Chretien, J. R.; Raevsky, O. A. Estimation of Blood-Brain Barrier Crossing of Drugs Using Molecular Size and Shape, and H-bonding Descriptors. *J. Drug Target.* **1998**, *6*, 151−165. In their experience, the conformational effect in low molecular weight molecules represents a variation of about ±10% in the polar surface area which they do not consider as critical when making first estimates of BBB uptake.

Acknowledgments

I am indebted to my colleagues from Rhône-Poulenc Rorer for many helpful discussions during the work described in this paper. In particular, Drs. Stephen Pickett, Andrew Ratcliffe, and Christopher Newton gave much encouragement and made constructive comments upon the manuscript while it was in preparation. Drs. Nicola Wilsher, Martin Barrett, and Ken Page gave advice on matters relating to absorption while Drs. Richard Lewis (now Eli Lilly, UK) and Paul Bamborough (now GlaxoWellcome, UK) helped in the development of the intranet interface to the calculations. I am grateful to Drs. Kristina Luthman (University of Tromsø, Norway) and Larry Dodd (Polytechnic University, New York) for the kind provision of values for atomic radii (Table 4) and the MOLVOL program, respectively. Prof. Peter Jurs (Pennsylvania State University) and Dr. Inge Christensen (Royal Danish School of Pharmacy) are thanked for providing preprints of refs 19 and 27, respectively. Finally, the comments and suggestions of the anonymous referees were greatly appreciated; particularly regarding the history of the use of molecular surface areas in modeling of transport and partitioning processes (refs 21 and 22) and issues concerning Boltzmann averaging of molecular surface areas.

JS9804011