Influence of Molecular Flexibility and Polar Surface Area Metrics on Oral Bioavailability in the Rat

Jing J. Lu,[†] Kimberly Crimin,[‡] Jay T. Goodwin,[#] Patrizia Crivori,[||] Christian Orrenius,[⊥] Li Xing,[△] Peter J. Tandler,[●] Thomas J. Vidmar,[×] Benny M. Amore,[§] Alan G. E. Wilson,[◆] Pieter F. W. Stouten,[⊥] and Philip S. Burton*[,#]

Computational Chemistry and Nonclinical Statistics and CNS Drug Metabolism, Pfizer Global R&D, Groton Laboratories, Eastern Point Road, 8200-36, Groton, Connecticut 06340, ADMETRx, Inc., 4717 Campus Drive, Suite 600, Kalamazoo, Michigan 49008, Attrition Reducing Technologies and Computational Sciences, Nerviano Medical Science, Viale Pasteur 10, 20014 Nerviano (MI), Italy, Computational Chemistry, Pfizer Global Research & Development, 700 Chesterfield Parkway North, St. Louis, Missouri 63017, 128 S. Burdick Street, Kalamazoo, Michigan 49007, Midwest Nonclinical Statistics, Pfizer, Kalamazoo, Michigan 49007, and Lexicon Genetics, 8800 Technology Forest Place, The Woodlands, Texas 77381

Received December 31, 2003

The relationship of rotatable bond count (N_{rot}) and polar surface area (PSA) with oral bioavailability in rats was examined for 434 Pharmacia compounds and compared with an earlier report from Veber et al. (*J. Med. Chem.* **2002**, *45*, 2615). N_{rot} and PSA were calculated with QikProp or Cerius2. The resulting correlations depended on the calculation method and the therapeutic class within the data superset. These results underscore that such generalizations must be used with caution.

We read with interest the publication from GlaxoSmithKline (GSK), which reports significant influences of molecular flexibility and polar surface area on oral bioavailability of early phase drug discovery candidates from the SmithKline Beecham compound collection.[1] In that study it was found that molecules possessing fewer than 10 rotatable bonds and having a polar surface area less than 140 Å^2 after oral administration generally showed bioavailability in the rat exceeding 20%. These results suggested that such criteria could be used as filters in the early discovery setting to identify drug candidates with minimally acceptable oral bioavailability and have generated significant interest in such approaches.[2−7]

These finding were of particular interest to us because we have been skeptical of the ability to predict complex pharmacokinetic endpoints from simple molecular descriptors.[8] Oral bioavailability is dependent on a number of different properties such as drug solubility and permeability, formulation factors, and physiological factors including regional permeability differences in the gastrointestinal tract, pH differences, lumenal and mucosal enzymology, intestinal motility, and first-pass metabolism, among others. While some of these processes may be influenced by molecular flexibility and/or polar surface area, it seems unlikely that such relationships would be similar. In the case of flexibility, for example, increased rotational degrees of freedom may result in a larger diffusional cross section, thereby influencing permeability adversely.[9] On the other hand, such flexibility could decrease crystallinity of the solute, resulting in improved aqueous solubility and enhanced absorption.[10,11] Similar antagonistic structure−property relationships may be expected for other factors contributing to bioavailability in the rat.

Nevertheless, the findings of Veber et al.[1] were intriguing and suggested that such an empirical correlation might have utility in the early phase drug discovery setting. Prior to possible implementation and routine use as a filter, we decided to validate these observations with our own data. To do so, we collected data on 434 Pharmacia compounds, primarily from programs in infectious disease, central nervous system, oncology, and cardiovascular disease, along with smaller numbers from a variety of other programs. Each compound had been dosed both intravenously and orally in rats to calculate oral bioavailability. Dosing vehicles varied depending on the program team and properties (i.e., solubility and dose) of the solute. However, no attempts were made to try to normalize for such differences in experimental protocol.

Methods

For the calculation of molecular descriptors, multistructure 2D SD files were generated for the 434 compounds and converted to 3D structures using Concord.[12] The resulting multiconformer files were used as input for QikProp (version 1.6)[13] to generate a series of 2D and 3D solvation-relevant descriptors. No energy minimization of resulting structures was performed other than the relaxation of close intramolecular contacts within Concord. The 3D conformations thus obtained are consistent for all homologous structures; consequently any differences in descriptor values associated with conformation may be minimized.

QikProp was used to obtain rotatable bond count (QikProp descriptor "#rotor"; in this paper termed N_{rot}), defined as the

* To whom correspondence should be addressed. Phone: 269-372-3272. Fax: 269-372-3397. E-mail: psburton@admetrx.com.
[†] Computational Chemistry, Pfizer Global R&D, Groton Laboratories.
[‡] Nonclinical Statistics, Pfizer Global R&D, Groton Laboratories.
[#] ADMETRx, Inc.
[||] Attrition Reducing Technologies, Nerviano Medical Science.
[⊥] Computational Sciences, Nerviano Medical Science.
[△] Computational Chemistry, Pfizer Global Research & Development, MO.
[●] 128 S. Burdick Street.
[×] Midwest Nonclinical Statistics.
[§] CNS Drug Metabolism, Pfizer Global R&D, Groton Laboratories.
[◆] Lexicon Genetics.

10.1021/jm0306529 CCC: $27.50
Published on Web 10/23/2004

Table 1. Oral Rat Bioavailability Quartile and Molecular Weight Comparison of GSK (n = 1100) Data and Pharmacia Data Set (n = 434)

	quartile			
	1	2	3	4
% *F* GSK, range	0−4.3	4.3−15.5	15.5−42.7	>42.7
% *F* Pharmacia	0−5.3	5.3−17.9	17.9−45.6	>45.6
MW GSK, mean	511.1	492.3	483.9	431.6
MW Pharmacia	536.7	499.9	493.1	398.5

number of nontrivial (not CX3, like in ethane) and nonhindered (not alkene, amide, small ring) rotatable bonds.

Similarly, polar surface area (QikProp descriptor "FISA"; in this paper termed PSA) is defined as the hydrophilic component of the SASA (SASA on N, O, and H attached to heteroatoms). SASA is the total solvent-accessible surface area in Å^2 using a spherical probe of 1.4 Å radius.

The two-dimensional SD files were also used directly as input into Cerius2[14] for generation of rotatable bond count, defined as the number of bonds in the current molecule having rotations that are considered to be meaningful for molecular mechanics. All terminal H atoms are ignored (for example, methyl groups are not considered rotatable).

Topological polar surface area (TPSA) descriptors in Cerius2 are akin to those described by Ertl et al.,[15] utilizing connectivity information to generate a van der Waals based PSA without having to account for three-dimensional structure and variability due to conformational flexibility.

Statistical analyses were performed using JMP,[16] and figures were generated with Spotfire.[17]

Results and Discussion

Of the 434 compounds in our data set, total clearance was reported for 415 of them. The relationship between bioavailability and clearance for these compounds is shown in Supporting Information and is similar to that reported by Veber et al., with a more extensive data set.[1] Although the relationship is not strong, on average the higher bioavailability compounds showed lower clearances and conversely the highest clearance compounds had lower average bioavailability. However, the ranges of bioavailability for a given clearance value indicated that other factors have significant influence on the bioavailabilities of these compounds in the rat.

To explore the influence of rotatable bond count on bioavailability and to compare the results with the GSK findings, we divided our data set into bioavailability quartiles following the GSK protocol. The resulting quartile comparisons are shown in Table 1. In general, we see surprisingly similar bioavailability ranges and average molecular weights for our data set compared to the GSK report.

Other comparisons of our data set with that of Veber et al. are included in Supporting Information. Briefly, both molecular weight and % *F* distributions were similar. In contrast, while 50% of GSK compounds had ClogP less than about 4.2, the Pharmacia 50% value was 1.8, suggesting some differences in the structural characteristics of the two libraries.

Table 2 summarizes the rotatable bond number and polar surface area values for each of our bioavailability quartiles from both the QikProp and Cerius2 programs. Clearly these results are quite different, with Cerius2 yielding approximately 2.5 rotatable bonds more than QikProp and significantly lower polar surface areas. For the PSA results, QikProp estimates solvent-accessible surface area while Cerius2 measures the van der Waals

Table 2. Comparison of N_{rot} and PSA Calculated from QikProp and Cerius2 for Pharmacia Data Set (n = 434) with GSK Previously Reported Results

	bioavailability quartile			
	1	2	3	4
N_{rot} QikProp, mean	10.6	8.5	8.5	5.1
N_{rot} Cerius2, mean	12.9	11.0	10.9	7.6
PSA QikProp, mean (Å^2)	177.3	168.6	151.5	156.4
PSA Cerius2, mean (Å^2)	119.2	108.4	102.1	88.5
N_{rot} GSK, mean[a]	10.2	9.0	8.2	6.2
PSA GSK, mean[a] (Å^2)	123.2	103.6	94.0	87.8

[a] Data from ref 1.

surface, which is the major reason for the difference. The method employed by Veber et al. for estimating PSA is an atom-based calculation of the van der Waals surface.[15] Indeed, when we compared results for several molecules from Ertl et al.[15] and the Goodman and Gilman data set in ref 1, Cerius2 gave results most similar to the results of the atom-based method while QikProp PSA values were consistently larger. Because Cerius2 uses a methodology that comes closest to the one employed by the GSK authors, for the purposes of the present study, we used the Cerius2 PSA values. Comparing these average surface areas with those from GSK in Table 2, we observe similar results within the quartile distributions, as was the case with molecular weights in Table 1.

In the case of rotatable bond number, QikProp gave consistently lower counts than Cerius2 (Table 2). While there is no generally accepted definition for rotatable bonds, QikProp counts only bonds that when rotated lead to unique conformers, e.g., gauche and anti butane. Restriction of rotation about such bonds upon binding to a protein then leads to an entropy penalty. Consequently, rotation of methyl or CF_3 groups is not included in the count, but R−OH bond is included, for example. Further, ring bonds in under-16-membered rings and any rotations about multiple bonds and hindered single bond cases such as amide OC−N and ester and acid OC−O are excluded (W. L. Jorgensen, personal communication). These considerations seem to be similar to the criteria used by GSK as discussed in the manuscript.[1] While the source of the discrepancy between QikProp and Cerius2 rotatable bond counts is not completely clear, some insights may be gained by comparing counts for reference compounds. When QikProp and Cerius2 counts were compared for the Goodman and Gilman data set,[1] as expected, QikProp gave values essentially identical with values from Veber et al., while Cerius2 counts were consistently higher. In a comparison of several structures such as phenytoin and clozapine, differences in allowable ring rotations seem to be contributing to the differences. Similary, in the case of salicylic and acetyl salicylic acid, differences seem to be present in the rotations counted in the carboxylic acid group. In summary, given these differences in polar surface area estimation and rotatable bond count between QikProp and Cerius2, we confirmed that the combination of Cerius2 PSA and QikProp N_{rot} would be most directly comparable to the methods used by the GSK group.

Figure 1 shows the selection efficiency of the combination of 10 or fewer rotatable bonds from QikProp and PSA less than or equal to 140 Å^2 from Cerius2, com-

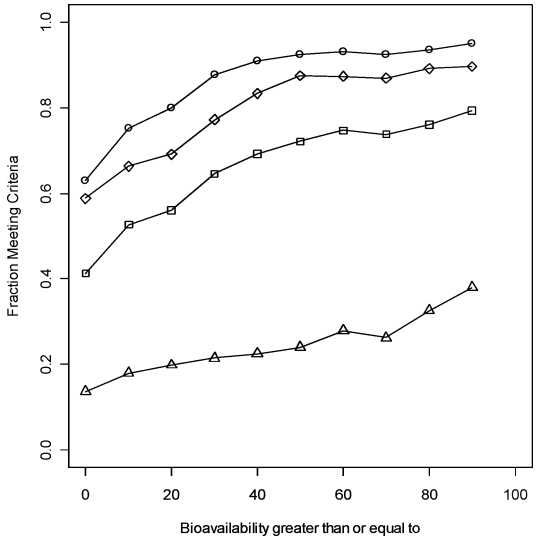

Figure 1. Efficiency of $N_{rot} \leq 10$ and PSA ≤ 140 Å^2 for selecting orally bioavailable compounds: (open circles) GSK results from Veber et al.;[1] (diamonds) Pharmacia data with QikProp N_{rot} and Cerius2 PSA; (squares) Pharmacia data with Cerius2 N_{rot} and Cerius2 PSA; (triangles) Pharmacia data with QikProp N_{rot} and QikProp PSA.

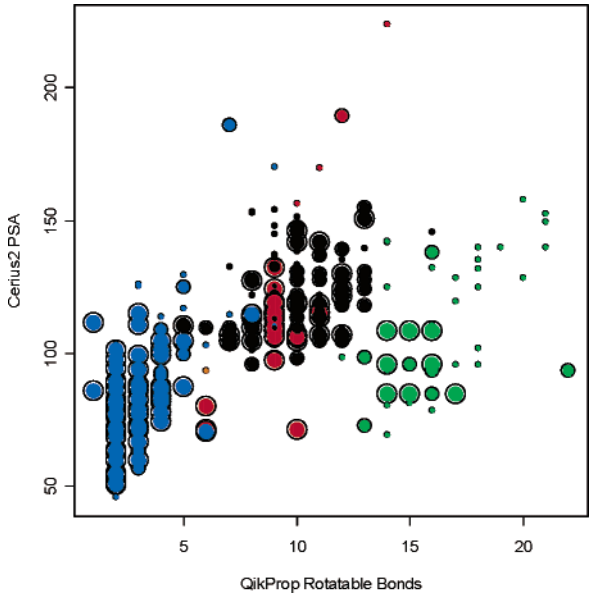

Figure 2. Rat oral bioavailability as a function of Cerius2 PSA and QikProp N_{rot} for Pharmacia compounds. The four different colors represent four different therapeutic areas. The bioavailability of each compound is represented by the size of its symbol (smallest symbol 0−20%; largest symbol 80−100%).

pared to the previously reported results from GSK. For our data set, 70% of compounds with bioavailability exceeding 20% meet these criteria. By comparison, GSK found 80% of their compounds meeting these criteria. Increasing the minimum bioavailability requirement results in a greater percentage of compounds meeting these flexibility and PSA constraints for both data sets, with our collection showing on average 5−10% more compounds with acceptable bioavailability not meeting the rotatable bond count and/or PSA filter. Included in Figure 1 are the results obtained from both N_{rot} and PSA from QikProp or Cerius2 for comparison. Clearly, in both these cases significantly fewer compounds are identified as acceptably bioavailable, especially in the case of QikProp. Interestingly, a comparison of the QikProp N_{rot}/QikProp PSA combination and Cerius2 N_{rot}/Cerius2 PSA with the QikProp N_{rot}/Cerius2 PSA combination shows the profound influence of PSA algorithm and less significant N_{rot} sensitivity. Collectively these results show that if such criteria are to be used as a filter, the choice of descriptor algorithm should be carefully validated with a relevant set of data.

A further consideration in the use of such filtering methods in the discovery setting is illustrated in Figure 2. In this figure we have identified the members of the major therapeutic subsets of compounds in our data superset and plotted them as a function of polar surface area and rotatable bond count. Also shown is the approximate bioavailability of each compound. The "privileged property space" of $N_{rot} \leq 10$ and PSA ≤ 140 Å^2 is in the lower left section of the plot. While we do see significant population in this region for structures from two project teams, another two projects are not so clearly defined, with one having essentially no members with these properties despite clearly acceptable oral bioavailability in the rat. It is particularly interesting that different chemical classes, representative of the unique therapeutic targets, cluster around different combinations of N_{rot} and PSA. Clearly, these properties are not homogeneously represented over our entire, albeit limited, chemistry space. Since the size of the symbols for each structural class represents the relative bioavailability, it is clear that within each cluster, compounds with acceptable bioavailability can be identified. This observation underscores the potential danger of attempting to generalize a very complicated endpoint and of using that generalization in a prospective selection application.

Supporting Information Available: Relationship of rat bioavailability with clearance for 415 Pharmacia compounds, property distributions, box plots comparing QikProp and Cerius2 PSA and N_{rot} calculations with GSK results for the 277-compound Goodman and Gilman human oral bioavailability data set, and the fraction of Pharmacia compounds with 20% oral bioavailability as a function of molecular weight and rotatable bond count. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615−2623.
(2) Muegge, I. Selection Criteria for Drug-Like Compounds. *Med. Res. Rev.* **2003**, *23*, 302−321.
(3) Abraham, M. H.; Ibrahim, A.; Zissimos, A. M.; Zhao, Y. H.; Comer, J.; Reynolds, D. P. Applications of Hydrogen Bonding Calculations in Property Based Drug Design. *Drug Discovery Today* **2002**, *7*, 1056−1063.
(4) Rishton, G. M. Nonleadlikeness and Leadlikeness in Biochemical Screening. *Drug Discovery Today* **2003**, *8*, 86−96.
(5) Bergstrom, C. A. S.; Strafford, M.; Lazarova, L.; Avdeef, A.; Luthman, K.; Artursson, P. Absorption Classification of Oral Drugs Based on Molecular Surface Properties. *J. Med. Chem.* **2003**, *46*, 558−570.
(6) Zmuidinavicius, D.; Didziapetris, R.; Japertas, P.; Avdeef, A.; Petrauskas, A. Classification of Structure−Activity Relationships (C-SAR) in Prediction of Human Intestinal Absorption. *J. Pharm. Sci.* **2003**, *92*, 621−633.
(7) Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. A Comparison of Physicochemical Property Profiles of Development and Marketed Oral Drugs. *J. Med. Chem.* **2003**, *46*, 1250−1256.

(8) Burton, P. S.; Goodwin, J. T.; Vidmar, T. J.; Amore, B. M. Predicting Drug Absorption: How Nature Made It a Difficult Problem. *J. Pharmacol. Exp. Ther.* **2002**, *303*, 889−895.
(9) Jackson, M. J. Drug Transport across Gastrointestinal Epithelia. In *Physiology of the Gastrointestinal Tract*, 2nd ed.; Johnson, L. R., Ed.; Raven Press: New York, 1987; pp 1597−1621.
(10) Vippagunta, S. R.; Brittain, H. G.; Grant, D. J. W. Crystalline Solids. *Adv. Drug Delivery Rev.* **2001**, *48*, 3−26.
(11) Bauer, J.; Spanton, R.; Henry, J.; Quick, J.; Dzidi, W.; Porter, W.; Morris, J. Ritonovir: An Extraordinary Example of Conformational Polymorphism. *Pharm. Res.* **2001**, *18*, 859−866.
(12) Concord (v.4.0.7; www.tripos.com).
(13) QikProp (v.1.6; see http://www.schrodinger.com/Products/qikprop.html).
(14) Cerius2 (www.accelerys.com/cerius2).
(15) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714−3717.
(16) JMP (www.jmp.com).
(17) Spotfire (www.spotfire.com).

JM0306529