Prediction of Blood-Brain Partitioning and Human Serum Albumin Binding Based on COSMO-RS σ-Moments

Karin Wichmann,*,† Michael Diedenhofen,† and Andreas Klamt†,‡

COSMO*logic* GmbH & Co. KG, Burscheider Strasse 515, 51381 Leverkusen, Germany, and Institute of Physical and Theoretical Chemistry, University of Regensburg, 93040 Regensburg, Germany

Received September 1, 2006

Models for the prediction of blood-brain partitioning (logBB) and human serum albumin binding (logK(HSA)) of neutral molecules were developed using the set of 5 COSMO-RS σ-moments as descriptors. These σ-moments have already been introduced earlier as a general descriptor set for partition coefficients. They are obtained from quantum chemical calculations using the continuum solvation model COSMO and a subsequent statistical decomposition of the resulting polarization charge densities. The model for blood-brain partitioning was built on a data set of 103 compounds and yielded a correlation coefficient of $r^2 = 0.71$ and an rms error of 0.40 log units. The human serum albumin binding model was built on a data set of 92 compounds and achieved an r^2 of 0.67 and an rms error of 0.33 log units. Both models were validated by leave-one-out cross-validation tests, which resulted in $q^2 = 0.68$ and a qms error of 0.42 for the logBB model and in $q^2 = 0.63$ and a qms error of 0.35 for the logK(HSA) model. Together with the previously published models for intestinal absorption and for drug solubility the presented two models complete the COSMO-RS based set of ADME prediction models.

INTRODUCTION

Within drug discovery, a large amount of failed compounds have problems related to their absorption, distribution, metabolism, or excretion (ADME) profile. ADME prediction is therefore becoming an increasingly important tool in the evaluation of compounds early in the drug discovery process. Experimental methods for determining properties like blood-brain partitioning are expensive and time-consuming. Hence predictive models for the evaluation of ADME properties, even before the compounds are synthesized, are important tools to enhance productivity in drug discovery.

In a complex biological system there exists a variety of influences on a drug. There are different mechanisms of actions for certain compounds, passive and active transportation (e.g., across the blood-brain barrier), multiple ways of binding (e.g., multiple binding sites in serum albumin), and various pathways of metabolism. Even if the complete life cycle of a drug in such a complex biological system cannot be modeled completely, it is generally preferred to model an individual property without explicitly considering other effects that might influence the property in question, while recognizing that this approach will introduce inaccuracies for some molecules.

A large number of QSAR based models for ADME property prediction methods have been published during the past decade. In silico ADME models for predicting blood-brain partitioning and serum albumin binding include correlation of experimental data with descriptors from Monte Carlo simulations of molecules in water,[1] polar surface area (PSA),[2] free energy of solvation from a generalized Born/surface area continuum solvation model,[3] VolSurf descriptors derived from 3D molecular fields,[4,5] topological structure descriptors,[6,7] and combinations of descriptors from quantum chemical calculations with complementary constitutional, topological, and chemical descriptors.[8,9]

Our special approach is the usage of the novel, quantum chemically based COSMO-RS approach for the prediction of ADME properties. The focus of our models is mainly on adsorption and distribution, while metabolism or excretion phenomena are not discussed. Because COSMO-RS does not only derive the information about the molecular interactions from quantum chemistry but also includes a rather rigorous treatment of the fluid phase thermodynamics, COSMO-RS based ADME models typically require less empirically adjusted parameters than other ADME prediction models and hence have the chance to be more predictive for drugs involving novel kinds of chemical functionality. COSMO-RS based models for two most important properties, i.e., for aqueous drug solubility and for intestinal absorption, have been published earlier.[10,11] This paper concerns the two remaining commonly considered ADME properties, i.e., blood-brain partitioning and drug−protein binding.

One of the first and often most important ADME properties to be considered during drug development is the potential of a drug molecule to penetrate the blood-brain barrier. Only substances targeting the central nervous (CNS) system should penetrate the blood-brain barrier, while other peripherally acting agents must exhibit low ability to pass the barrier. Drug entry into the brain is a complex process which depends on several factors. While there exist some ways of active transport by catalyzed transport systems for certain more polar compounds, the free (passive) diffusion of drugs across the endothelium separating the blood from the CNS is prevented by endothelial tight junctions of severly limited

* Corresponding author e-mail: wichmann@cosmologic.de.
† COSMO*logic* GmbH & Co. KG.
‡ University of Regensburg.

10.1021/ci600385w CCC: $37.00
Published on Web 12/29/2006

permeability. Consequently, the majority of compounds entering the brain by passive diffusion will use a transcellular route. Therefore we will consider the passive membrane permeation only in this study. Although blood-brain permeation is physicochemically much less cleanly defined than the commonly considered, but physiologically irrelevant, octanol−water partition coefficient, it is justified to consider the blood-brain partition coefficient as a generalized partitioning problem.

The other ADME property considered in this paper is the binding to human serum albumin. Drug−protein binding affects the pharmacological activities and side effects of drugs as well as drug distribution and elimination. Serum albumin is the most abundant protein in plasma and the major transport medium for nonesterified fatty acids as well as for drugs and metabolites. The availability of a drug to a target tissue is affected by its binding strength to serum albumin. The reversible drug−protein complex serves as a depot making unbound drug available when the concentration of the free drug decreases. The free concentration of the drug, hence the biological activity, depends on the amount of drug bound to serum proteins. The consequences of this may be prolonged activity, which might be desirable or not, or the emergence of undesirable side effects. Treating the albumin part of the blood as a pseudoseparate phase, the human serum albumin binding coefficient can rather obviously be considered as a generalized partition coefficient.

Theoretical Framework. The COSMO-RS method, a combination of the quantum chemical continuum solvation model (COSMO) and a statistical thermodynamics treatment for real solvents (RS) simulations, is a widely applicable tool for accurate predictions of many kinds of thermodynamic and physiological fluid phase properties. Since a full derivation of the theory of COSMO-RS is beyond the scope of this article, only a short summary of the essentials is given here. More details can be found in refs 10 and 12−17.

COSMO-RS considers a liquid system as an ensemble of molecules, including the pure solvent or a solvent mixture and one or more solutes. A precondition for COSMO-RS is a DFT/COSMO[12] calculation for each compound X in the ensemble, to get the total energy E^X_{COSMO} and the screening charge density (SCD) σ on the molecular surface. σ is a very good local descriptor for molecular surface polarity. In the COSMO-RS theory, all surfaces of the molecules in the ensemble are then assumed to be in pairwise close contact. The electrostatic interaction of the screening charges σ and σ' on a surface pair is described by

$$E_{misfit}(\sigma,\sigma') = \frac{\alpha'}{2}(\sigma + \sigma')^2 \quad (1)$$

Hydrogen-bonding energy can also be approximately quantified by the two adjacent SCDs. Since only hydrogen bond donors have a very strong negative SCD and only hydrogen bond acceptors have a strong positive SCD, a functional of the form

$$E_{HB}(\sigma,\sigma') = c_{HB}\min\{0,\sigma\sigma' + \sigma^2_{HB}\} \quad (2)$$

can be used to describe hydrogen bonding (HB) energies. The parameters c_{HB} and σ_{HB} have been appropriately adjusted in the COSMO-RS parametrization. This simple functional already expresses that more polar donors and acceptors make stronger hydrogen bonds than less polar ones.

In addition to electrostatic misfit and hydrogen bonding interactions, COSMO-RS also takes into account van der Waals (vdW) interactions between surface segments via

$$E_{vdW}(e,e') = \tau(e) + \tau(e') \quad (3)$$

where e and e' indicate the elements of the atoms below the surface segments and $\tau(e)$ is an adjustable element specific parameter. But since vdW interactions are much less specific than polar interactions and hydrogen bonds and only become important in the transfer of a solute to the gas phase, this part is not relevant for the calculation of partition coefficients of solutes between liquid phases. Thus for the following we may assume that all interactions are just pairwise interactions of SCDs σ and σ'.

The transition from microscopic molecular surface charge interactions to macroscopic thermodynamic properties of mixtures is made via a statistical thermodynamic procedure. Since only pair wise interactions have to be considered in the statistical ensemble S, it turns out that the statistical averaging can be done on an ensemble of interacting surface pieces. The ensemble is fully characterized by the screening charge distributions $p^X(\sigma)$ of the different molecular species X in S, in the following called σ-profiles. The σ-profile of a mixture $p_S(\sigma)$ is built additively from the σ-profiles of the compounds weighted with their mole fraction. The chemical potential $\mu_S(\sigma)$ of a surface segment with SCD σ in an ensemble described by the distribution $p_S(\sigma)$ is

$$\mu_S(\sigma) = -\frac{RT}{a_{eff}}\ln\left\{\int p_S(\sigma')\exp\left(\frac{a_{eff}}{RT}(\mu_S(\sigma') - E(\sigma,\sigma'))\right)d\sigma'\right\} \quad (4)$$

This implicit equation, in which a_{eff} denotes an effective, statistically independent piece of contact area and $E(\sigma,\sigma')$ denotes the sum of the energy contributions from eqs 1 and 2, can be solved by iteration within milliseconds on a PC. The function $\mu_S(\sigma)$ describes the affinity of a solvent S for surfaces with polarity σ, and it will be called σ-potential further on. The chemical potential of a solute X in the solvent S is now mainly given by a surface integral of the σ-potential $\mu_S(\sigma)$ over the surface of the solute X:

$$\mu^X_S = \int p^X(\sigma)\mu_S(\sigma)d\sigma + \mu^X_{S,comb} \quad (5)$$

The combinatorial contribution $\mu^X_{S,comb}$ which takes into account size effects of solvents and solutes, usually is of small importance and can be neglected for the partition properties considered in this paper. Equation 5 gives access to the entire fluid phase equilibrium thermodynamics of almost arbitrary compounds in liquid systems, including all kinds of partition coefficients between well characterized liquid phases as for example the octanol−water partition coefficient. It has been widely validated versus most different experimental solubility and partition data.

Nevertheless, several important thermodynamic and especially physiologic equilibria involve one or more phases, which are either chemically less well defined, or which are disordered, but not really liquid, or both. Examples of such

systems are physiological phases like blood, brain, the intestinal resorption system, or other special tissue, environmental systems as soil, or technically important systems as polymers and solid adsorbents, like activated carbon. In such phases no exact chemical compositon of the effective solvent phase and hence no surface composition function $p_S(\sigma)$ is available. Hence the σ-potential $\mu_S(\sigma)$ of the phase S and the chemical potentials $\mu^X{}_S$ of solutes X in these phases cannot be directly calculated by COSMO-RS. However, an indirect treatment of such phases by COSMO-RS is enabled by the σ-moment extension, which can be motivated by the following consideration: Experience from a large number of different solvents led us to the finding that σ-potentials $\mu_S(\sigma)$ can be described very accurately by a Taylor-like expansion of the form

$$\mu_S(\sigma) \cong \sum_{i=-2}^{m} c_S^i f_i(\sigma) \tag{6}$$

with

$$f_i(\sigma) = \sigma^i \quad \text{for} \quad i \geq 0 \tag{7}$$

$$f_{-2}(\sigma) = f_{acc}(\sigma) \cong \begin{cases} 0 & \text{if } \sigma \leq \sigma_{HB} \\ \sigma_{HB} - \sigma & \text{if } \sigma > \sigma_{HB} \end{cases} \tag{8a}$$

$$f_{-1}(\sigma) = f_{don}(\sigma) \cong \begin{cases} 0 & \text{if } -\sigma \leq \sigma_{HB} \\ \sigma_{HB} + \sigma & \text{if } -\sigma > \sigma_{HB} \end{cases} \tag{8b}$$

The highest order required for a sufficient description typically is $m = 3$. In this sense, we may characterize each solvent (at fixed temperature, usually room temperature) by the set of σ-coefficients $c^i{}_S$. Obviously any difference between the σ-potentials of two solvents is of the same kind of expansion, with coefficients $c^i{}_{S,S'}$ being just the difference of the coefficients of the two solvents. As a consequence, any logarithmic partition coefficient between two phases S and S′ (e.g., aqueous solution and membrane) should then be expressible as a linear combination of the drug σ-moments, in the form

$$\log K^X_{S,S'} = \tilde{c}_{S,S'} + \int p^X(\sigma) \sum_{i=-2}^{m} \tilde{c}^i_{S,S'} f_i(\sigma) d\sigma$$

$$= \tilde{c}_{S,S'} + \sum_{i=-2}^{m} \tilde{c}^i_{S,S'} M^X_i \tag{9}$$

where all coefficients $\tilde{c}_{S,S'}$ depend only on the two solvent phases, and the σ-moments $M_i{}^X$ of the solute X are defined by

$$M^X_i = \int p^X(\sigma) f_i(\sigma) d\sigma \tag{10}$$

The σ-moment approach is very versatile and has been successfully applied to such different problems as intestinal absorption,[10] adsorption onto activated carbon,[18] and soil sorption coefficients.[19] It has also been shown that the σ-moments are a rather complete set of descriptors for partitioning, which is valuable in any QSAR study involving partition behavior. Although derived from a theoretical approach, they are strongly related to the empirically derived set of the 5 Abraham LFER descriptors, which have been widely and successfully used in the description of partitioning and ADME properties since 1990.[20] It should be noted that most of the 5 σ-moments have a simple and well-defined meaning: The zeroth-order moment is just the COSMO surface area of the solute, and the first-order moment is the negative of the total charge and hence vanishes for all partitioning data sets which usually take into account neutral compounds only. The second-order moment can be shown to correspond to the negative of the electrostatic interaction energy of the solute with a dielectric continuum, while the third-order moment does not have a simple interpretation. The acceptor and donor moments resulting from the acceptor and donor functions ($i = -2,-1$, respectively) defined in eq 8, quantify the ability of the solute to interact as acceptor and donor, respectively. A more detailed description of the σ-moment approach is given in a recent review.[10]

METHODS

Experimental Data. Data for logBB were taken from a study of Rose et al.[6] In the original paper 103 data points were used for the training set, all of which have also been used in this study. The data set consists of neutral compounds, including a wide range of molecular size and complexity.

Data for logK(HSA) were taken from a study of Colmenarejo et al.[21] The data were derived from HPLC chromatographic retention time for immobilized human serum albumin for a set of 94 drugs.

Computational Procedure. 3D structures of molecules were built using CORINA.[22] Molecular geometries were optimized using the AM1 semiempirical method and the Conductor-like Screening Model (COSMO), as implemented in the MOPAC 2002 program,[23] using optimized COSMO radii.[24,25] The AM1 geometries were used to do single point density functional calculations to produce COSMO files holding the surface polarization charges. DFT/COSMO calculations were carried out using the BP functional, the SVP basis set, and the RI-DFT method, employing the TURBOMOLE 5.6 program package.[26] The COSMO*therm* software was used to calculate the σ-moments.[27]

According to the motivation given in the previous section our σ-moment models for logBB and logK(HSA) were constructed by linear regression of the logarithmic partition properties vs the set of the 5 σ-moments of order -2, ...,3, omitting the first order σ-moment, which vanished for our sets of neutral solutes.

RESULTS AND DISCUSSION

σ-Moment models for logBB and logK(HSA) were calculated from multilinear regression. The models were internally validated by leave-one-out cross-validation.

Blood-Brain Partitioning. The logBB regression model and statistics are

$$\log BB = 0.0046(\pm 0.0006)M_0 - 0.0173(\pm 0.0012)M_2 - 0.0027(\pm 0.0021)M_3 + 0.1878(\pm 0.1048)$$

$$n = 103,\ r^2 = 0.71,\ \text{rms} = 0.40,\ q^2 = 0.68,\ \text{qms} = 0.42 \tag{11}$$

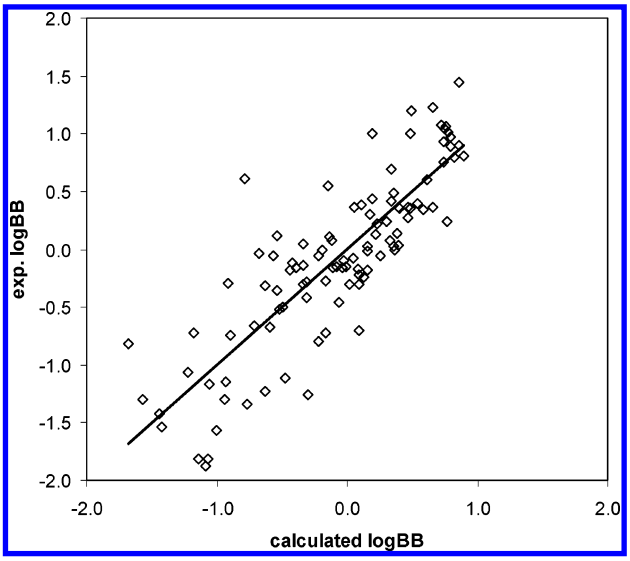

Figure 1. Experimental vs calculated logBB.

Here n denotes the numbers of compounds in the training sets, r^2 is the squared correlation coefficient, and the rms is the root-mean-square error. q^2 and qms are the corresponding predictive correlation coefficients and errors based on leave-one-out cross-validation.

The logBB model (Figure 1) employs only three descriptors, M_0, M_2, and M_3, since the hydrogen bond donor and acceptor descriptors M_{don} and M_{acc} turn out to be statistically insignificant for the logBB model. Apparently, the descriptors M_2 (overall electrostatic interaction) and M_3 capture the correlation of the blood-brain partitioning with the molecular polarity sufficiently. In most QSAR studies of blood-brain partitioning, either a desciptor of the molecular polarity *or* a descriptor of the hydrogen bond capacity is found to correlate negatively with logBB; our findings suggest that it is rather the overall polarity of a compound than its hydrogen bonding ability that leads to reduced brain permeation. The positive sign of the zero-order (surface area) coefficient and the negative sign of the second order (electrostatic) coefficient describe the commonly known trends that unpolar molecules cross the barrier more easily than polar molecules.

The logBB model yields a correlation coefficient r^2 of 0.71 and an rms of 0.40. In view of the range of experimental logBB values from approximately −2 to 1.5 log units this seems a considerable error. However, in a recent review, Clark[28] compared a variety of logBB models and found that the predictive performance of all of them is about the same with ∼0.4 log units, which appears to be close to the experimental error of ∼0.3 log units. The accuracy of the in silico methods is apparently limited by experimental noise in the logBB measurements. The strongest deviation of the predicted logBB from the experimental data is found for oxazepam. It is noteworthy that oxazepam is also among the strongest outliers in the studies of Rose et al.[6] and Hutter.[8]

The logBB model in eq 11 has been derived from molecular conformations as generated by the standard procedure described above. In order to investigate the effect of conformations on the model, we built a second logBB model, where several conformers per compound were generated on the AM1 gas-phase level of theory, and the lowest energy conformation of these was chosen. BP/SVP COSMO single point calculations were performed, and from the resulting COSMO files another set of σ-moment descriptors was derived to build the second logBB model. Coefficients and statistics of this model are very similar to those of the previously described procedure, with $r^2 = 0.71$, rms = 0.40, $q^2 = 0.68$, and qms = 0.42. From this we can conclude that the effect of the molecular conformation is of minor importance, at least for logBB.

There are several QSAR studies of blood-brain partitioning using the same or similar data sets. Our σ-moment logBB model can be compared directly to the model of Rose et al.[6] The five topological structure descriptors employed in their study include the hydrogen bond donation ability, a descriptor for nonpolar aromatic CH groups, a measure for the degree of branching in the structure, a generalized measure of nonpolarity, and a composite index for electron accessibility from flourine and chlorine atoms and the number of the two halogens. With statistical values of $r^2 = 0.73$, $s = 0.40$, $q^2 = 0.70$, and s_{press} = qms = 0.43, this model performs similarly to our σ-moments logBB model.

A data set of 85 compounds, with substantial overlap with the data set used in this study, was used by Kaznessis et al.[1] to develop three logBB models. The descriptors used are properties like the solvent-accessible surface area (SASA), the numbers of hydrogen bond donors and acceptors, the solute dipole, and the hydrophilic, hydrophobic, and amphipilic components of the SASA, as determined from Monte Carlo simulations of the compounds in water. In each model, several severe outliers were removed. The resulting statistical quantities for the three different models are $r^2 = 0.97$ and $s = 0.173$, $r^2 = 0.958$ and $s = 0.203$, and $r^2 = 0.932$ and $s = 0.256$, respectively.

The model obtained by Ooms et al.[4] was built on a data set of 83 compounds, which also has a large overlap with the data set used in this study. Their final model includes 31 independent VolSurf descriptors of polarity, hydrophobic interaction, and shape and size. After removing four outliers, it gives an r^2 of 0.76 and a q^2 of 0.65.

Keserü et al.[3] developed a logBB model based on the correlation of solvation free energies and blood-brain partitioning, using a smaller data set of 55 compounds. The solvation free energy G_{solv} is the only descriptor, and the model yields an r of 0.85 ($r^2 = 0.72$) and a standard error of 0.37.

Another logBB model is obtained from a training set of 90 descriptors by Hutter.[8] The final model uses of 12 of the 41 computed descriptors, some of them derived from semiempirical AM1 calculations, others from the molecular structures or 3D geometries of the compounds. On the training set, the model gives an r^2 of 0.865 and a standard error of estimate of 0.309.

Human Serum Albumin Binding. The human serum albumin binding model, logK(HSA), obtained by multilinear regression from the σ-moments is

$$\log \mathrm{K(HSA)} = 0.0084(\pm 0.0007)M_0 - 0.0166(\pm 0.0015)M_2 - 0.0138(\pm 0.0022)M_3 + 0.1645(\pm 0.0336)M_{acc} - 1.0201(\pm 0.1536)$$

$$n = 94,\ r^2 = 0.67,\ \mathrm{rms} = 0.33,\ q^2 = 0.63,\ \mathrm{qms} = 0.35 \quad (12)$$

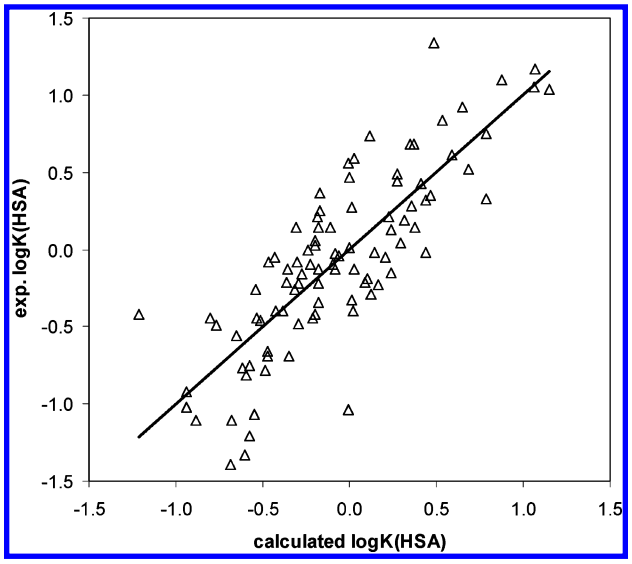

Figure 2. Experimental vs calculated logK(HSA).

This model (Figure 2) employs four of the five σ-moment descriptors, M_0, M_2, M_3, and M_{acc}. The hydrogen bond donor descriptor was omitted due to statistical insignificance, suggesting that the hydrogen bond donor ability is not important for human serum albumin binding. The descriptor for hydrogen bond acceptor ability M_{acc} seems to be of some importance and correlates positively with logK(HSA). M_2 and M_3 correlate inversely with logK(HSA). The model yields an r^2 of 0.67 and an rms of 0.33, lower than the statistical quantities of the logBB model. The largest deviation of the predicted logK(HSA) from the experimental data is found for ebselen with 1.03 log units. Ebselen is also among the compounds with the largest residuals in the study of Kier et al.[6] although with a smaller deviation from experimental data than in our study.

The logK(HSA) model from this study is built on the data set that was also used in two previous studies. Thus, we can compare directly to the models from these studies. Since in both studies the data set was divided into a training set of 84 compounds and a test set of 10 compounds, we recalculated the model for this smaller training set. The resulting logK(HSA) model achieves a correlation coefficient r^2 of 0.66 and an rms 0.33, a q^2 of 0.61, and a qms of 0.36. For the external validation test set, the correlation coefficient is $r^2 = 0.80$, and the rms error is 0.32. The largest residual is found for acetylsalicylic acid with 0.75 log units.

Colmenarejo et al.[9] optimized a logK(HSA) model by a genetic algorithm, resulting in a nonlinear equation with 6 quantum chemical, topological, and chemical descriptors. The model yields an r^2 of 0.83 and a $q^2 = 0.79$ on the training set and an r^2 of 0.82 on the external validation set. The largest deviation was also found for acetylsalicylic acid.

Kier et al.[6] also used the same data set to model the binding of drugs and druglike compounds to human serum albumin. They found a model with 6 topological descriptors with $r^2 = 0.77$, $s = 0.29$, $q^2 = 0.70$, and $s_{press} = 0.33$, which are slightly better statistical quantities than for our σ-moment logK(HSA) model. In the external test set, the correlation coefficient is $r^2 = 0.74$ and the mean absolute error MAE $= 0.31$. The largest outlier in the external validation is acetylsalicylic acid, as in the study by Colmenarejo et al. and in this study.

CONCLUSION

σ-Moment based QSAR models are broadly applicable to the a priori prediction of a variety of partitioning questions. The σ-moment approximation of the COSMO-RS method is theoretically well justified and has been successfully applied to a large variety of chemical and physiological partition properties involving phases of less well-defined composition. While being based on a consistent and theoretically justified small set of descriptors, the predictive qualities of the logBB and the logK(HSA) regression models presented in this study are about the same as those of most previously published in silico prediction models for these properties. The accuracy of all of these models can be considered as limited by the experimental noise of the available experimental data sets.

The suite of the previously published COSMO-RS models for aqueous drug solubility and intestinal absorption, together with the new models for blood-brain partitioning and human serum albumin binding, now offers a consistent set of in silico prediction models for the most widely considered ADME properties in drug design. σ-Moment models for additional properties can easily be added on demand, following the same framework as described above. An advantage of the σ-moment approach is the completeness of the descriptor set, which renders the evaluation of a large amount of possibly relevant yet correlated descriptors unnecessary. Another advantage is the rather fundamental, quantum chemical definition of the σ-moments, which makes σ-moment models applicable to almost all kinds of organic compounds, even if they should involve rare chemical functionality. As a possible disadvantage one might consider the time-demanding nature of the quantum chemical COSMO calculation. However, this step which usually only takes ~10 min per drug compound, has to be done only once per molecule in order to have the entire range of σ-moment descriptors and hence all COSMO-RS based properties available. Furthermore, for high-throughput tasks it can be substituted by a suitable fragmentation into molecules from a precalculated database using the COSMO*frag* methodology,[29] with only a very small loss of accuracy.

ACKNOWLEDGMENT

We thank Lowell H. Hall and L. Mark Hall for supplying structures and tables of data in electronic format.

Supporting Information Available: Experimental and calculated data used in this study (Tables 1 and 2). This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Kaznessis, Y. N.; Snow, M. E.; Blankley, J. C. Prediction of blood-brain partitioning using Monte-Carlo simulations of molecules in water. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 697−708.

(2) Ertl, P.; Rohde, B.; Selzer, P. Fast contribution of molecular polar surface area as a sum of fragment-based contributions and its applications to the prediction of drug transport. *J. Med. Chem.* **2000**, *43*, 3714−3717.

(3) Keserü, G. M.; Molnár, L. High-Throughput Prediction of Blood-Brain Partitioning: A Thermodynamic Approach. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 120−128.

(4) Ooms, F.; Weber, P.; Carrupt, P.-A.; Testa, B. A simple model to predict blood-brain barrier permeation from 3D molecular fields. *Biochim. Biophys. Acta* **2002**, *1587*, 118−125.

(5) Aureli, L.; Cruciani, G.; Candida Cesta, M.; Anacardio, R.; De Simone, L.; Moriconi, A. Predicting Human Serum Albumin Affinity of Interleukin-8 (CXCL8) Inhibitors by 3D-QSPR Approach. *J. Med. Chem.* **2005**, *48*, 2469−2479.

(6) Rose, K.; Hall, L. H.; Hall, L. M.; Kier, L. B. Modeling Blood-Brain Barrier Partitioning Using Topological Structure Descriptors, 2003. Elsevier MDL Web Site. http://www.mdl.com/solutions/white_papers/MDLQSARreprint.jsp (accessed Feb 13, 2004).

(7) Kier, L. B.; Hall, L. H.; Hall, L. M. QSAR Modeling of Drug Binding to Protein. β-Lactam Serum Binding and Albumin Binding Affinity, 2002. Elsevier MDL Web Site. http://www.mdl.com/solutions/white_papers/ML233QM3_white_paper.jsp (accessed Oct 7, 2004).

(8) Hutter, M. C. Prediction of blood-brain barrier permeation using quantum chemically derived information. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 415−433.

(9) Colmenarejo, G.; Alvarez-Pedraglio, A.; Lavandera, J.-L. Cheminformatic Models to Predict Binding Affinities to Human Serum Albumin. *J. Med. Chem.* **2001**, *44*, 4370−4378.

(10) Klamt, A.; Eckert, F.; Hornig, M. COSMO-RS: A novel view to physiological solvation and partition questions. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 355−365.

(11) Jones, R.; Connolly, P. C.; Klamt, A.; Diedenhofen, M. Use of Surface Charges from DFT Calculations To Predict Intestinal Absorption. *J. Chem. Inf. Model.* **2005**, *45*, 1337−1342.

(12) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799−805.

(13) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224−2235.

(14) Klamt, A. COSMO and COSMO-RS. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, L., Eds.; Wiley: New York, 1998; Vol. 2, pp 604−615.

(15) Klamt, A.; Jonas, V.; Buerger, T.; Lohrenz, J. C. W. Refinement and Parameterization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.

(16) Klamt, A.; Eckert, F. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilib.* **2000**, *172*, 43−72.

(17) Klamt, A. *COSMO-RS. From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier: Amsterdam, The Netherlands, 2005; pp 49−147.

(18) Mehler, C.; Klamt, A.; Peukert, W. Use of COSMO-RS for the prediction of adsorption equilibria. *AICHE J.* **2002**, *48*, 1093−1099.

(19) Klamt, A.; Eckert, F.; Diedenhofen, M. Prediction of Soil Sorption Coefficients with COSMO-RS. *Environ. Toxicol. Chem.* **2002**, *21*, 2562−2566.

(20) Zissimos, A. M.; Abraham, M. H.; Klamt, A.; Eckert, F.; Wood, J. A Comparison between the Two General Sets of Linear Free Energy Descriptors of Abraham and Klamt. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1320−1331.

(21) Colmenarejo, G.; Alvarez-Pedraglio, A.; Lavandera, J.-L. Cheminformatic Models to Predict Binding Affinities to Human Serum Albumin. *J. Med. Chem.* **2001**, *44*, 4370−4378.

(22) Sadowski, J.; Gasteiger, J. From atoms and bonds to three-dimensional atomic coordinates: automatic model builders. *Chem. Rev.* **1993**, *93*, 2567−2581.

(23) *MOPAC 2002*, version 2.00; Fujitsu: Tokyo, Japan, 2002.

(24) Klamt, A.; Eckert, F. Fast Solvent Screening via Quantum Chemistry: The COSMO-RS approach. *AIChE J.* **2002**, *48*, 369−385.

(25) For Selen a recently optimized COSMO radius of 2.200 Å was used.

(26) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165−169.

(27) *COSMOtherm*, version C2.0 release 01.04; COSMO*logic* GmbH & Co. KG: Leverkusen, Germany, 2004.

(28) Clark, D. E. In-silico prediction of blood-brain barrier permeation. *Drug Discovery Today* **2003**, *8*, 927−933.

(29) Hornig, M.; Klamt, A. COSMO*frag*: A Novel Tool for High Throughput ADME Property Prediction and Similarity Screening Based on Quantum Chemistry. *J. Chem. Inf. Model.* **2005**, *45*, 1169−1177.

CI600385W