Pharmaceutical Research, Vol. 26, No. 1, January 2009 (© 2008)
DOI: 10.1007/s11095-008-9733-x

Research Paper

Development of Improved Empirical Models for Estimating the Binding Constant of a β-Cyclodextrin Inclusion Complex

Ravi Chari,[1] Farooq Qureshi,[2] John Moschera,[2] Ralph Tarantino,[2] and Devendra Kalonia[1,3]

Received July 16, 2008; accepted September 19, 2008; published online October 9, 2008

Purpose. To develop empirical models for predicting the binding between a drug and β-cyclodextrin. Specifically, the logarithm of the 1:1 binding constant is expressed as the function of various molecular descriptors of the drug. Many potential drugs exhibit poor aqueous solubility. Also, the amount available for solubility studies is limited early in drug development. Thus, models that show which excipients can increase a drug's solubility are useful because formulation scientists can focus on them experimentally.
Methods. Twenty-five descriptors were considered based on molecular characteristics governing complexation. These include the drug's size and/or shape, the dispersion of its electron cloud, its lipophilicity, and its flexibility. The training set contains 258 ligands, ranging from drug-like molecules to small polar organic compounds.
Results. Two models were developed. The first is derived by partial least squares regression and consists of all 25 descriptors. The r^2 determined by cross-validation is 0.79. The second contains four variables and was constructed by multiple linear regression. Its cross-validated r^2 is 0.65.
Conclusions. Due to its simplicity, the second model is recommended over the first. The most important descriptor in both models is the calculated log P, indicating that drugs with greater lipophilicity form stronger complexes with β-cyclodextrin.

KEY WORDS: binding; cyclodextrins; descriptors; empirical; models.

INTRODUCTION

Many drug candidates exhibit high lipophilicities and low water solubilities. Such characteristics typically correspond to poor bioavailability which reduces the likelihood of developing a successful therapeutic formulation. Therefore, techniques that increase the aqueous solubilities of such candidates are highly desired. One method involves their complexation with cyclodextrins (CDs), cylindrical excipients made up of glucose units (1–5). The most common unmodified CDs are α, β, and γ-CD which are composed of six, seven, and eight glucose units respectively. In order to increase their solubilities, a random number of their hydroxyl groups are substituted with various chemical moieties, including sulfobutyl ether or hydroxypropyl groups. This random substitution imparts an amorphous nature to the CDs which results in increased solubility (6). Nevertheless, the unmodified CDs still possess higher solubilities than the vast majority of hydrophobic drug candidates. CDs possess a torus-like structure in which each glucose unit is connected to an adjacent unit by 1,4 α-linkages (3, 4). The last unit is joined to the first, completing the ring. The conformation of CD resembles a bottomless bucket with both narrow and wider openings and an internal cavity. The primary hydroxyl groups are oriented towards the narrow entrance while the secondary hydroxyl groups face the wider opening (4). The potential for CDs to form inclusion complexes has been attributed to many factors, one of which is the lowered hydrophilicity of the internal cavity (3). Water molecules tend to exclude themselves from this region due to the higher energy required to extricate themselves from the hydrogen bonding network of the aqueous environment prior to entering the cavity (3). The lowered presence of water in the CD interior leads to an environment more

Molecular descriptor abbreviations are listed in Table I.

[1] Department of Pharmaceutical Sciences, School of Pharmacy, University of Connecticut, 69 North Eagleville Rd., U-3092, Storrs, CT 06269, USA.
[2] Hoffmann-La Roche Inc., Nutley, NJ, USA.
[3] To whom correspondence should be addressed. (e-mail: kalonia@uconn.edu)

ABBREVIATIONS: A_L, linear curve type; A_N, negative deviation from linearity; A_P, positive deviation from linearity; B, curve type where drug solubility decreases at high cyclodextrin concentrations; CD, cyclodextrin; [CD], cyclodextrin molar concentration; $\Delta G°$, standard free energy (of binding); K, 1:1 binding constant; Ln K, natural logarithm of the 1:1 binding constant; Log K, logarithm of the 1:1 binding constant; log P, logarithm of the octanol/water partition coefficient; MLR, multiple linear regression; NH, amino group; OH, hydroxyl group; p, probability value; PC, principal components; PLS, partial least squares; PLSR, partial least squares regression; R, gas constant; r, correlation coefficient; r^2, square of correlation coefficient; $r^2_{adjusted}$, square of adjusted correlation coefficient; r^2_{cv}, square of correlation coefficient from leave-one-out cross validation; S, total molar solubility; S_0, water solubility; SH, thiol group; T, temperature in Kelvin.

hydrophobic than that of bulk water and provides a pocket in which lipophilic compounds may reside (3). Such compounds may also be stabilized within the cavity by hydrogen bonding and/or the interaction of dispersion forces between the two species (3). The result, in ideal formulations, is a water soluble complex in which the CD is essentially a carrier of the drug compound. CDs can accommodate a wide range of drug sizes, ranging from 300 to 700 g/mol. Examples include nimesulide, a non-steroidal anti-inflammatory drug, complexed with β-CD (7,8). The drug has a molecular weight of 308.31 g/mol and a dissolution time of over two hours. Complexation with β-CD decreases that time to one hour. In contrast, itraconazole which is used to treat esophageal candidiosis has been co-formulated with hydroxypropyl β-CD (8,9). It has a much greater molecular weight of 705.71 g/mol and a solubility of less than 0.001 mg/ml. Its complexation with that CD increases its overall solubility to 17 mg/ml, an amount suitable for IV infusion. Thorough reviews on CD-drug formulations published in the last five years include those of Challa *et al.* (1), Loftsson and Duchene (2), and Szejtli (8). Other such reviews older than five years include those of Connors (3), Uekama *et al.* (4), Stella and Rajewski (5), and Loftsson (10).

While many mechanistic aspects of CD mediated drug solubilization can be explored, the purpose of this study was to develop an empirical model to predict the binding constant between a compound and β-CD based upon the values of certain molecular descriptors of the compound. Not only would such a model lead to a better understanding of the drug characteristics important for complexation, it would also serve as an important tool for facilitating drug formulation, especially with regard to poorly soluble compounds. At the early stages of drug development, the amount of drug available for carrying out solubility studies with various excipients is usually insufficient. Therefore, the availability of a model to predict whether a particular excipient or technique is feasible for increasing the solubility of a potential drug is highly desired. In addition to preserving the limited amount of drug available, time, labor, and the cost of materials are also reduced. The empirical models developed here predict the binding constant between a drug and β-CD. The value determines whether the drug is a good candidate for inclusion within this CD and therefore whether the CD may be used to increase the drug's solubility to the dosage required for administration. This is based upon the inference that the value of the binding constant is directly related to the amount of drug-CD complex present and therefore the amount of drug solubilized by the CD. A review of the literature concerning complexation or solubilization of drug-like molecules with CDs reveals a number of different empirical models. Although most show high predictive ability, there are concerns with their utility and whether their results can be confidently accepted. For example, the small size of the training set of a three term model by Suzuki *et al.* (11) may limit the predictive ability of that model. The narrow range of compounds in the training sets of two models by Klein *et al.* (12, 13) also has the same effect. Finally, the fragment based approach utilized by Suzuki *et al.* (14) and Katritzky *et al.* (15) restricts the types of compound that can be analyzed. These models are compared and contrasted with the models developed in this study in the "DISCUSSION" section.

MATERIALS AND METHODS

Summary

Our models predict the 1:1 binding constant (K) between a drug and β-CD based upon the values of various molecular descriptors of the drug. The descriptors were selected based on a literature review of the characteristics important for complexation. The training set used to derive the models was assembled by mining the scientific literature for studies on β-CD complexation. The experimental binding constants from these studies are for the 1:1 association between drug and CD. Finally, both partial least squares regression and multiple linear regression were used to derive the models.

Molecular Descriptors: Calculation and Selection Rationale

The compounds of the training set were constructed with the molecular modeling software package HyperChem v7.51 (Hypercube Inc., Gainesville, FL USA) and were saved in that program's HIN format. The software Dragon Professional v5.4 was then used to calculate the compounds' molecular descriptor values (16). The 25 descriptors are listed in Table I and were evaluated for model inclusion based upon their incorporation in previous models from the literature as well as on their putative importance in explaining the interactions between the compounds and β-CD. In addition, the values of the descriptors are not dependent on molecular conformation. Some of the training set compounds have many possible conformations and considering them all is beyond the scope of this study. The following four paragraphs summarize the salient features of each descriptor while more thorough descriptions can be found in the comprehensive *Handbook of Molecular Descriptors* by Todeschini and Consonni (17).

Although many instances of β-CD complexation are enthalpically driven, the hydrophobic effect still plays a major role in the association of compounds with the CD (18). This is supported by the observation that the more non-polar regions of a ligand typically partition into the relatively hydrophobic CD interior while the remaining polar areas face the aqueous environment (3). Therefore descriptors of hydrophobicity, including calculated log P (ALOGP), the hydrophilic factor (HY), and topological polar surface area (TPSA), were considered for the model. ALOGP is determined by the Ghose–Crippen method where a molecule's atoms are classified according to a set of 120 pre-defined atom types (17,19). The assigned hydrophobic constants of those atoms are summed to yield the log P value. Higher values correspond to increased lipophilicity. HY quantitates the hydrophilicity of a molecule based upon the number of hydrophilic groups (–OH, –SH, –NH). Higher values indicate greater hydrophilicity. TPSA is the surface area of a molecule's polar atoms—oxygen, nitrogen, and sulfur as well as any attached hydrogen atoms—determined by a fragment based approach (20).

The other factor affecting complexation is the compound's size and shape. For example, a compound may be too large or too small to fit optimally within a CD cavity. It may also be too spherical to complement the CD's toroid shape or, conversely, its degree of ellipticity may be adequate for a

Table I. Molecular Descriptors Selected for Constructing the Empirical Models

Symbol	Name	Property	Training set average	Training set range
ALOGP	Ghose-Crippen octanol–water partition coeff. (log *P*)	Hydrophobicity/ hydrophilicity	1.9±1.2	−2.0–5.7
HY	Hydrophilic factor	Hydrophobicity/ hydrophilicity	0.03±0.80	−0.9–3.1
TPSA	Topological polar surface area using N,O,S,P polar contributions	Hydrophobicity/ hydrophilicity	46±37	0–182.8
MW	Molecular weight	Size	199±109	32–765
SV	Sum of atomic van der Waals volumes (scaled on Carbon atom)	Size	16±9	3–66
SPH	spherosity	Shape	0.8±0.2	0–1
S1κ	1-path Kier alpha-modified shape index	Shape	10±6	2–40
S2κ	2-path Kier alpha-modified shape index	Shape	4±2	1–15
S3κ	3-path Kier alpha-modified shape index	Shape	2.6±1.6	0–9
χ^0	Connectivity index chi-0	Shape	10±5	2–38
χ^1	Connectivity index chi-1 (Randic connectivity index)	Shape	6.5±3.7	1–26
χ^2	Connectivity index chi-2	Shape	5.9±3.8	0–26
χ^3	Connectivity index chi-3	Shape	4.6±3.6	0–23
χ^4	Connectivity index chi-4	Shape	3.6±3.2	0–19
χ^5	Connectivity index chi-5	Shape	2.7±2.7	0–15
J	Balaban distance connectivity index	Shape	2.1±0.4	0.8–3.5
SP	Sum of atomic polarizabilities (scaled on Carbon atom)	Electron cloud interactions	17±10	3–71
SS	Sum of Kier–Hall electrotopological states	Electron cloud interactions	35±20	8–122
AMR	Ghose–Crippen molar refractivity	Electron cloud interactions	53±28	8.3–191.8
AROM	Aromaticity index	Electron cloud interactions	0.7±0.5	0–1
JGT	Global topological charge index	Electron cloud interactions	0.47±0.15	0–0.77
nCIR	Number of circuits	Flexibility	2.4±3.1	0–15
nCIC	Number of rings	Flexibility	1.6±1.3	0–8
PHI	Kier flexibility index	Flexibility	3.1±1.9	0.6–11.5
RBF	Rotatable bond fraction	Flexibility	0.07±0.06	0–0.24

complementary match. In addition, the amount of branching can have a dual effect. A certain degree of branching may be necessary to achieve optimal van der Waals contacts with the CD interior while excess branching may lead to steric clashes between the compound and the CD interior or even prevent binding by hindering the compound's entry. Descriptors of size include the molecular weight (MW) and van der Waals volume (SV) while shape descriptors include spherosity (SPH) and the Kier–Hall α-modified shape indices (S1κ, S2κ, and S3κ). SPH indicates the degree to which a molecule's shape resembles a sphere with greater values corresponding to a less spherical and more planar form. S1κ describes the relative cyclicity of a molecule while S2κ measures the spatial density or branching. S3κ provides information on the centrality of branching, *i.e.* whether branching is more prevalent at the extremities or at the center of a molecule. The connectivity indices (χ^0, χ^1, χ^2, χ^3, χ^4, and χ^5) were also considered in developing the model. χ^0 is dependent on the degree of saturation or branching of a compound while χ^1 and χ^2 quantitate the degree of (non-double or -triple bond) saturation about single and double bond units, respectively (11). χ^3 to χ^5 provide similar, but unique, information regarding more complex bond units. Finally, the Balaban distance-sum connectivity index (*J*) was also evaluated. *J* measures the number of bond units that each atom is apart from every other atom in a compound. Higher values indicate smaller distances, reflecting a molecule's compactness. Within an isometric series, highly branched isomers possess greater *J* values than more linear forms.

A third factor affecting complexation is the interaction between electron clouds of a compound and those of β-CD (3). The resulting polarization leads to stabilization of the

complex through complementary van der Waal contacts between the two species. The magnitude of such interactions correlates with the size of the compound because larger molecules possess greater electron clouds. Therefore, many descriptors of electron density will also have a component reflecting compound size. Such descriptors include the sum of atomic polarizabilities (SP), the sum of Kier Hall electrotopological states (SS), and the molar refractivity (AMR). SP reflects the hardness or the degree to which a molecule can be polarized, as opposed to the degree to which it is polarized. SS sums the electronic accessibilities of a molecule's atoms and represents their probability of interaction with another molecule. AMR is a measure of the volume occupied by an atom or molecule. Closely related to spatial density, it quantitates the amount of matter filling that volume. Two additional descriptors, the aromatic index (AROM) and the global topological charge index (JGT), were also considered. Unlike the previous three, both are independent of molecular size. AROM simply reflects the presence of aromatic groups while JGT provides information on charge transfer in a molecule as it relates to topology.

Finally, the fourth factor influencing complexation is the flexibility or rigidity of a compound. This characteristic reflects a compound's ability to adapt to the shape of the CD cavity (12). Thus a large, but flexible, molecule can deform in order to achieve a complementary fit, whereas a rigid compound of the same size would be unable to bind. Descriptors that measure flexibility include the number of rings (nCIC) or circuits (nCIR) in a molecule, the Kier molecular flexibility index (PHI), and the fraction of freely rotatable bonds over all bonds present (RBF). nCIC and nCIR serve as measures of rigidity with higher numbers of rings or circuits corresponding to reduced flexibility. PHI is proportional to the number of rotatable bonds and is inversely proportional to the amount of rings present. Higher values of both PHI and RBF indicate greater flexibility.

Training Set Characteristics

The training set consists of 258 observations of unique β-CD–ligand complexes and was constructed by mining the scientific literature for studies on β-CD–ligand interactions. Some studies focused on the binding of a particular compound or group of compounds with β-CD (18,21–26) while others detailed the development of other empirical models to predict β-CD – ligand complexation (11,12,27–31). Many of the studies determined the 1:1 binding constants from phase solubility curves, the majority of which were found to be linear (A_L-type) (12). Others used titration calorimetry to determine the binding constants (21). Yet others measured the binding of a ligand based upon how well it displaced a reference ligand from the β-CD interior (31). It should be noted that the latter two techniques do not yield phase solubility curves, and thus the curve type could not be obtained. Fortunately, this is not an issue since this study is not concerned with the prediction of higher-order complex formation. In addition, knowledge of the phase solubility curve type does not affect the procedures used in this study. Only the 1:1 β-CD–ligand binding constant is needed. These values were recorded in the training set. If a study reported binding free energies instead, they were converted to binding constants via the equation $\Delta G^\circ = -RT \ln K$ where ΔG° is the Gibbs standard free energy of binding, R is the gas constant, T is the temperature in units of Kelvin, and Ln K is the natural logarithm of the 1:1 binding constant.

All studies were performed at 25°C in water. No co-solvents were added, although salts or buffers have been used in some in order to regulate the pH, and therefore control the ionization state of acidic or basic ligands. Of the 258 ligands in the training set, 112 were found to be ionizable, *i.e.* consisting of amines, carboxylic acids, phenolic OH groups, or combinations of the three. The majority of these (~90%) were studied under pH conditions at which they were neutral.

Many of the training set compounds are either drugs or possess drug-like structures while others are small organic molecules. The latter consists of both non-polar and polar compounds, with some having high miscibility in water. The molecular descriptors of the compounds were calculated as described above and entered into the training set. The average molecular weight (MW) of the compounds is 199± 109 g/mol and ranges from 32 to 765. The average log P (ALOGP) is 1.9±1.2 and ranges from −2.0 to 5.7. Table I shows the averages and ranges of all 25 descriptors for all compounds of the training set.

CD–Ligand Database

A database was developed with Microsoft Office Access 2003 (Microsoft Corp., Redmond, WA USA) in order to store information compiled from various CD–ligand binding studies. The training set for this study was generated by screening it for β-CD–ligand binding experiments performed at 25°C in water. The architectural format of the database consists of three separate tables. The first contains entries for 13 different CDs and includes fields for properties such as CD name, number of sugar units, substituents, and the degree of substitution. The second table has entries for 350 ligands with fields for the 25 molecular descriptors previously mentioned as well as fields for ligand water solubility. In addition, both tables contain the Kekule structures of the corresponding molecular species. The third table consists of 642 unique entries, each containing the experimental conditions and results from studies on a specific CD–ligand pair. The table is dynamically linked to the previous two tables in order to facilitate the entry of new data. For example, if a new study contains a CD which has been previously entered into the first table, the user can simply establish a relationship between the entry for that CD in the first table and a new entry in the third table. This way, the CD characteristics need not be reentered each time that excipient appears in a binding or solubility study with a unique ligand. Fields in the third table include those for experimental conditions such as temperature, pH, ionic strength, buffers, solvents, co-solvents, co-solutes, and CD and ligand formal charges. Fields for experimental results contain binding constants for different CD-ligand stoichiometric ratios as well as binding free energies. Finally, a search form based upon a global search query containing the fields from all three tables was set up to allow for quick retrieval of the information from those separate tables on a single form.

Statistical Techniques

Minitab 14.1 (Minitab Inc., State College, PA USA) was used to perform the statistical analyses, including multiple linear regression, partial least squares regression, principal component analysis, and the comparison of correlation coefficients. Microsoft Office Excel 2003 (Microsoft Corp., Redmond, WA USA) was used to organize the data and facilitate the analysis.

RESULTS

β-CD Partial Least Squares Model

The partial least squares (PLS) model predicts the Log *K* of binding between a compound and β-CD based on the values of the compound's molecular descriptors. It was initially constructed using multiple linear regression and contained all 25 descriptors. However, the variance inflation factors indicated that many of the terms were highly collinear and therefore partial least squares regression (PLSR) was employed to build this model.

PLSR is an extension of principal component analysis (32,33). The technique redefines the coordinate space of the current variables resulting in an equal number of new variables, or principal components, that are based upon the variance in the data. The first principal component encompasses as much of the variance as possible with subsequent components spanning less of the remaining variance than previous ones. Collectively, the total of the principal components represents 100% of the variance. By including those principal components needed to maximize the predictive ability, or cross-validated r^2 (r^2_{cv}), in the model, PLSR has the advantage of addressing collinearity and noise without the need for variable reduction. In particular, collinearity is addressed by extracting specific principal components. A single principal component can be extracted from a group of highly collinear variables to represent the majority of the variance from that group. Noise from the data is addressed by utilizing only those principal components needed for maximizing r^2_{cv} . Including more principal components in the model will result in overfitting, *i.e.* the model will begin to predict the contribution from the noise.

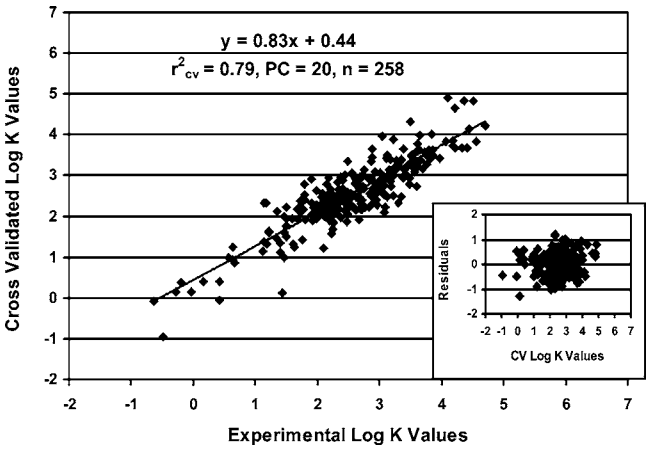

Fig. 1. Plot of Log *K* values predicted by the PLS model *versus* the experimental Log *K* values for the compounds of the training set. The predicted Log *K* values were determined by leave-one-out cross validation. The insert shows the residual plot. r^2_{cv} cross-validated r^2; *PC* principal components, *n* training set size.

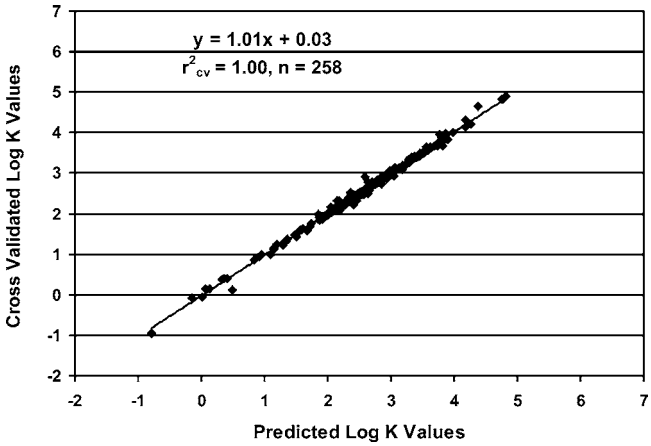

Fig. 2. For the PLS model, plot of Log *K* values predicted by leave-one-out cross-validation *versus* those predicted by implementing the full training set.

The PLS model was internally validated by leave-one-out cross-validation which resulted in an optimized model of 20 principal components with an r^2_{cv} of 0.79. Table II lists the regression coefficients while Fig. 1 shows the plot of cross-validated predicted Log *K* values *versus* the corresponding experimentally determined values for the compounds of the training set. Cross-validated predicted values are determined by omitting a single observation from the training set and calculating the predicted Log *K* for that observation based on the remaining training set. This is contrary to ordinary predicted values which are obtained without removing those observations from the training set. For the model, a plot of cross-validated predicted values *versus* ordinary predicted values shows that the correlation between the two is very high (Fig. 2). This shows that the training set size is large enough such that omission of a single observation does not affect its predictive ability. Fig. 1 also shows a plot of the residuals *versus* the experimental Log *K* values. Most of the residuals are below one Log *K* unit and show no apparent trend, indicating that non-linear or higher order terms in the model's equation are not needed.

The importance of each descriptor to the model was evaluated by ranking the magnitudes of the standardized coefficients (Table II). This is because studying the values of the regular coefficients may lead to the erroneous conclusion that a descriptor with a larger coefficient is more important to the model than one with a coefficient that approaches zero. For example, the coefficient of JGT is 1.501 while that of MW is 0.008, suggesting that JGT has greater influence. These coefficients, however, do not account for the differences in the magnitudes between the two descriptors' values. The average value of JGT for all compounds in the training set is 0.47±0.15 and ranges from 0 to 0.77. For MW the average is larger, 199±109, and ranges from 32 to 765. It is now clear that the small value of the regular coefficient of MW mitigates the large values of MW observed in the training set compounds. Conversely, the opposite applies to JGT.

Table II. Coefficients and Standardized Coefficients of the PLS Model

Descriptor symbol	Log *K* coefficient	Standardized coefficient
S1κ	−0.833	−5.115
SV	0.280	2.834
SP	0.238	2.553
χ^1	−0.603	−2.422
χ^2	0.512	2.129
S2κ	0.791	1.949
AMR	−0.059	−1.815
χ^3	0.439	1.730
χ^0	−0.254	−1.545
χ^5	−0.497	−1.440
MW	0.008	0.970
SS	0.038	0.825
nCIC	−0.581	−0.820
PHI	−0.328	−0.667
nCIR	0.182	0.614
ALOGP	0.438	0.584
χ^4	0.158	0.545
AROM	1.054	0.524
TPSA	0.009	0.369
RBF	4.276	0.274
SPH	−1.235	−0.243
JGT	1.501	0.239
J	0.497	0.202
S3κ	−0.088	−0.158
HY	−0.004	−0.003
Constant/Y-int	0.066	0

Thus, one must examine the standardized coefficients to account for this. They are obtained by transforming the set of values of each descriptor to possess a mean of zero and a variance of one. The standardized coefficients for MW and JGT are 0.970 and 0.239, respectively, indicating that MW is actually more important than JGT to the model. In other words, a change of one standard deviation of MW influences the predicted Log *K* to a greater degree than a one standard deviation change in JGT.

Ranking the standardized coefficients shows that S1κ is the most important descriptor in model. S1κ is the Kier–Hall α-modified shape index which is a measure of the relative cyclicity of a compound. A decrease in the value indicates an increase in cyclicity with multi-cyclic compounds having lower values than monocyclic ones. Since the S1κ coefficient is negative, an increase in S1κ yields a lower Log *K*, meaning that a decrease in cyclicity leads to lowered binding affinity. The sensitivity of Log *K* to changes in the cyclicity of a compound may be due to the cylindrical shape of the CD cavity. Circularly shaped compounds or compounds with circular moieties would better complement the CD interior than irregularly shaped compounds. The least important descriptor is the hydrophilic factor. HY is relatively insensitive to hydrophobic molecules, and since much of the training set consists of such compounds, the value of HY is not expected to change appreciably.

Additionally, the importance of the descriptors was also evaluated by removing each one from the model and determining the r^2_{cv}. This revealed that the calculated log *P* (ALOGP) is the most crucial by far. Its removal resulted in a much greater reduction in r^2_{cv} than the removal of any other descriptor, including S1κ. In contrast to HY, which is unimportant according to its low standardized coefficient, ALOGP is more sensitive to differences in hydrophobicity among the compounds of the training set. The descriptor's importance suggests that molecules with greater lipophilicity are more likely to partition into the relatively hydrophobic CD cavity. The concern exists, however, that since the calculated log *P* contains information on molecular size, the binding constant may depend on this aspect rather than on hydrophobicity (3). Fortunately, this is not the case since ALOGP is poorly correlated ($r<0.31$ at most) with all the other descriptors of the compounds of the training set, including size descriptors such as MW or SV.

β-CD Reduced Variable Model

To complement the PLS model, a simplified model was developed through a multi-step process involving the application of multiple linear regression (MLR). A flow chart of the procedure is shown in Fig. 3. First, the application of stepwise regression (α-to-enter and α-to-remove both set to 0.15) reduced the number of descriptors from 25 to 15. For this model, multicollinearity was then assessed by a combination of principal component analysis, correlation coefficient examination, and variance inflation factor evaluation. From a set of collinear descriptors, the descriptor whose values showed the highest correlation with the experimental Log *K* values was included in the model and the remaining discarded. This step left ten descriptors for further evaluation. Next, five non-significant descriptors were identified by *p* values >0.05 obtained through MLR and subsequently removed. In addition, the descriptor SV was replaced by SP. The two are highly correlated ($r=0.99$), but SP contains additional information on atomic polarization in addition to shared information on molecular size. Furthermore, RBF was removed. Like PHI, it is a measure of molecular flexibility but is the less significant term. The resulting model determines the Log *K* from the descriptors ALOGP, PHI, *J*, and SP. It was internally validated by leave-one-out cross-validation and has an r^2 of 0.65, an $r^2_{adjusted}$ of 0.64, and an r^2_{cv} of 0.63. The term r^2 is the square of the correlation coefficient. It is a measure of the overall variation between values predicted by the model *versus* experimentally determined values. The term $r^2_{adjusted}$ is a modified r^2 that accounts for the number of terms in the model. Adding unnecessary terms may result in a higher r^2 solely due to noise. The $r^2_{adjusted}$ accounts for this and may even decrease when such terms are included. Table III lists the regression coefficients while Fig. 4 shows the plot of the cross-validated predicted Log *K* values *versus* the experimental values for the compounds of the training set, as well as the residuals plot. Similar to the PLS model, most of the residuals are within one Log *K* unit. Their uniform distribution and lack of an apparent trend indicates that non-linear or higher order terms are not required. As with the PLS model, ALOGP was found to be the most important descriptor in the reduced variable model. Its absence resulted in a lower r^2_{cv} (0.36) than when any other descriptor was removed. In contrast, the removal of SP, PHI, or *J* yielded an r^2_{cv} of 0.43, 0.58, and 0.58, respectively.

The four terms that comprise the reduced variable model each quantitate a unique molecular characteristic and can be

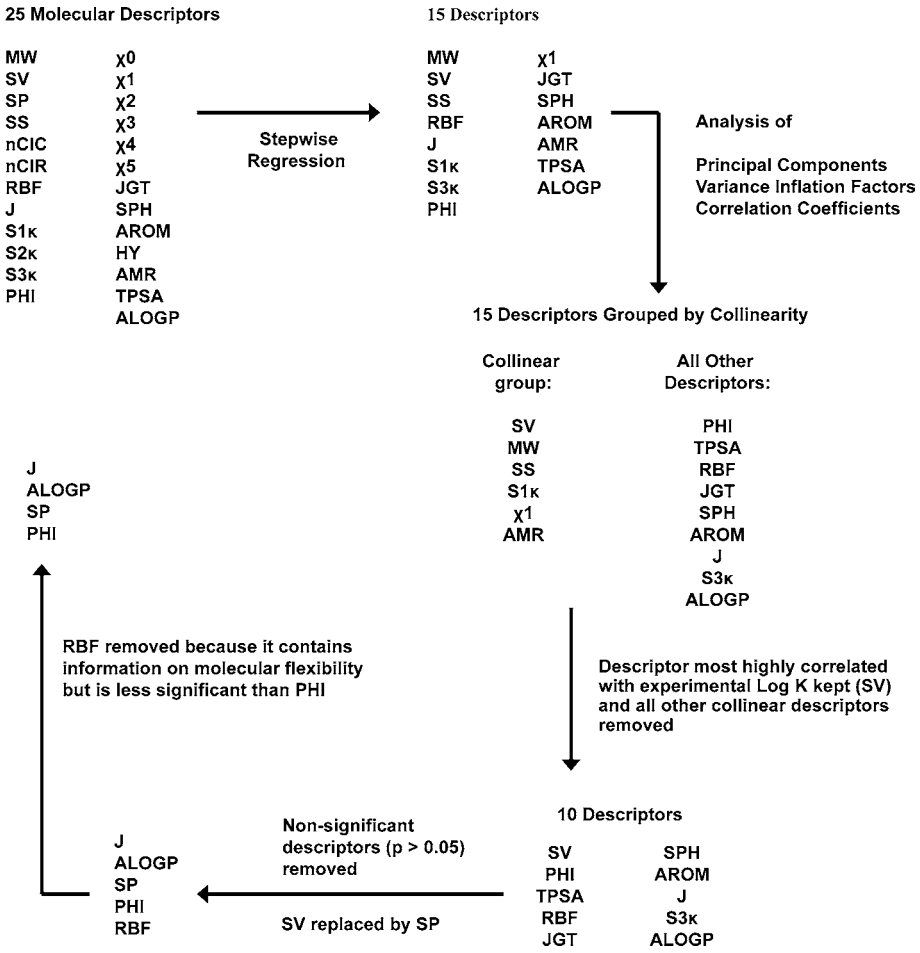

Fig. 3. Procedure used to reduce the number of terms to obtain the reduced variable model.

categorized in terms of thermodynamic principles. SP is a measure of polarizability or hardness and contains information on size. It reflects the steric effects of complexation and is thus related to the enthalpy of binding. The positive sign of the descriptor's coefficient suggests that a greater potential for polarization leads to an increase in binding. If a molecule's electron cloud is soft, it may be able to adjust in order to fit within the CD cavity. ALOGP conveys information on solvent entropy, *i.e.* the hydrophobic effect. As previously mentioned, the positive sign of ALOGP's coefficient shows that ligands with greater lipophilicity have a greater tendency to partition into the CD interior. The

Table III. Coefficients, Standardized Coefficients, and Descriptor *p* Values for the Reduced Variable Model

Descriptor symbol	Log *K* coefficient	Standardized coefficient	*p* values
Constant	−0.635	0	0.034
SP	0.077	0.829	<0.009
ALOGP	0.406	0.541	<0.009
PHI	−0.180	−0.366	<0.009
J	0.753	0.307	<0.009

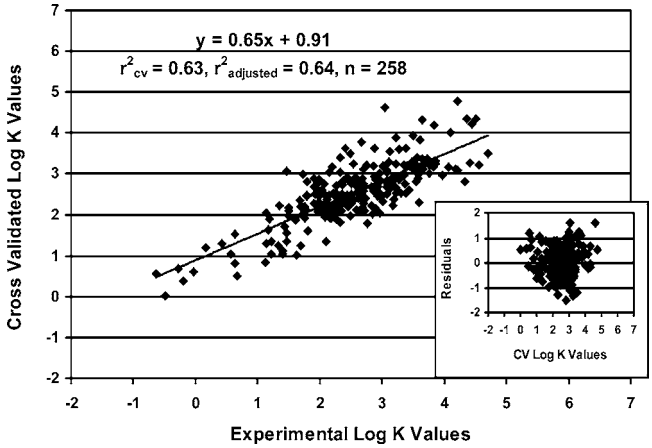

Fig. 4. Plot of the predicted Log *K* values determined by the reduced variable model *versus* the experimental Log *K* values for the compounds of the training set. The predicted Log *K* values were determined by leave-one-out cross validation. The insert is the residual plot. r^2_{cv} cross-validated r^2, *n* training set size.

descriptor PHI measures molecular flexibility which serves as an indicator of configurational entropy. Its negative coefficient indicates that binding decreases as the flexibility of a compound increases. Interestingly, this contrasts with the results from molecular modeling studies which show that an increase in configurational entropy, and presumably flexibility, actually leads to an increase in binding (34). Finally, *J* provides information on the effect of ligand shape or branching on complexation and, like SP, is related to the enthalpy of binding. The positive sign of its coefficient shows that an increase in compactness and/or branching leads to greater complex formation. This is not surprising since greater branching and more compactness ultimately results in a more spherical structure. Since the CD cavity is a cylinder, such a shape is an ideal fit.

DISCUSSION

Evaluation of the Training Set

The training set was assembled from molecules included in the training sets of previous β-CD complexation models as well as from compounds from various studies on β-CD–ligand association. Since we did not perform these studies ourselves, this precludes any firsthand knowledge of the researchers' proficiency and the exact techniques and equipment used to obtain the binding constants. Although the methods detailed in those publications suggest that experimental conditions were largely uniform—temperature set at 25°C and studies performed in water—the conditions are not exactly the same across all studies. As previously mentioned, in some studies the pH was varied in order to study the ionizable ligands in the neutral state. If the pH was not monitored carefully, the ionization state may change, thus affecting the binding constant. Indeed, studies have shown that the charged state of a ligand results in a lower binding constant than the neutral state (22). In addition, adjusting the pH requires the use of buffers which in turn affects the solution ionic strength. Some studies have found that an increase in ionic strength from ~0 to 900 mM has a negligible effect on CD-ligand complexation (35). Others, however, have observed a slight increase in binding from 10 to 300 mM ionic strength (36). The manifestation of such discrepancies and/or variations is not surprising when aggregating many separate studies. The only way to address these issues is to compile as large a training set as possible with the expectation that variations in the quality of the data will even out allowing for the true underlying trends to show through.

In addition, the method used to determine the β-CD–ligand binding constants was not used as a criterion to include or exclude compounds from the training set. It was observed, however, that the phase solubility method was most often used to determine that property (37). In these cases, the type of solubility curves found (A_P, A_L, A_N, or *B*) (37) was often not reported, although A_L-type curves predominated when it was mentioned. In addition, using the phase solubility method to determine the β-CD–ligand binding constant also requires the water solubility of the ligand. This value can be difficult to determine accurately, especially in the case of lipophilic drugs, and any resulting experimental error will carry over to the binding constant value. Other techniques used to determine the binding constants include titration calorimetry (21) and the measurement of a ligand's ability to displace a reference ligand from the β-CD interior (31). Since each technique has its own unique sources of error, and because the training set is essentially a compilation of these techniques, this supports the need for as large a training set as possible in order to smooth out such variations.

Furthermore, the training set was also constructed with as wide a range of molecules as possible in order to reduce the possibility of evaluating an underrepresented compound. A diverse training set will more likely ensure that the descriptor values of a novel compound will not fall at the extremes of the ranges of the descriptor values of the training set compounds.

Evaluation and Use of the Models

The PLS model is based on 25 descriptors and has an r^2_{cv} of 0.79 while the reduced variable model consists of four descriptors and has an r^2_{cv} of 0.65. This suggests that 21 descriptors essentially account for the 0.14 difference between the r^2_{cv} values of the two models. The observation that four descriptors constitute most of the variance may lead to concerns of both overfitting, *i.e.* modeling the noise, and inclusion of too many descriptors in the PLS model. It is for these reasons that PLSR was employed. Twenty principal components that contributed to the predictive ability were included in the model and six, based on the noise, were removed. In addition, concerns about excess descriptors are addressed by the rule-of-thumb that at least five compounds per variable are needed for a reliable model (38,39). The training set contains 258 molecules which is over ten compounds per descriptor.

Of the two, the reduced variable model is recommended for predicting the Log *K* of novel compounds. Although the PLS model shows better predictive ability, and the task of calculating its many descriptors can easily be performed by a simple computer program, the model may be influenced by overfitting. In other words, there may be concerns that some noise might be been incorporated despite having been optimized at 20 principal components. Thus, the reduced variable model is recommended because one can be certain that due to having only four descriptors, its predicted binding constants are largely free from noise. The concern in this case, however, is that this model has lost some predictive ability along with the noise. Nevertheless, this is mitigated by the likelihood that the model is more inclusive than the PLS model. Drugs with descriptor values that fall outside the range of those values of the compounds of the training set cannot be evaluated with confidence by the models. Such outlying values are more likely to be present when using the PLS model simply because it has many more descriptors than the reduced variable model.

A ligand is appropriate for evaluation by the models if its descriptor values lie within the range of values found in the training set (fifth column in Table I). It is preferable though if the ligand's descriptor values are within one standard deviation of the average descriptor values of the training set (fourth column in Table I). As mentioned, this is more likely when using the reduced variable model simply because it has less descriptors.

In addition to the binding constant, the models may be used to predict the solubility enhancement of a compound by

β-CD. This would estimate of the amount of β-CD needed in a formulation in order solubilize the required dosage of a drug. Assuming an A_L-type solubility curve, the total drug solubility can be determined with the Higuchi and Connors (37) equation

$$S = S_0 + \frac{KS_0}{1 + KS_0}[\mathrm{CD}]$$

where S is the total molar solubility of the drug, S_0 is the drug water solubility, K is the 1:1 β-CD–ligand binding constant, and [CD] is the total concentration of β-CD. If S_0 is not known it must be confidently estimated. In addition, [CD] must be less than 0.00163 M, the upper limit of β-CD's solubility. This implies that the total drug solubility cannot be greater than 0.00163 M, assuming a negligible S_0 and a 1:1 binding of CD and drug. One concern with estimating solubility this way is that many β-CD–ligand complexes are either insoluble or show decreasing solubility after a certain CD concentration (1). This behavior is reflected in B-type phase solubility curves. Unfortunately, the models developed here are not able to predict this possibility at the current time. Future work will focus on determining the ligand characteristics that govern it.

Comparison with Previous Models from the Scientific Literature

One of the goals of this study was to improve on the limitations of previous empirical models. For example, a three term model by Suzuki *et al.* was based on a training set of 33 diverse ligands ranging from drug-like molecules to small polar organic compounds (11). It has an r^2 of 0.92 and an r^2_{cv} of 0.90 and is as follows: $Log\ K = -0.095(\chi^0)^2 + 1.30(\chi^0) + 0.43 \log P - 2.85$. While the two r^2 values suggest good predictive ability, a plot of the Log K values predicted by the model *versus* the experimental values of our training set compounds shows a poor correlation (r^2=0.07) for many of the ligands (Fig. 5). Further analysis attributed the poor fits to the χ^0 term. The values of that descriptor in the training set compiled by Suzuki *et al.* ranges from zero to 9.5 which is much narrower than that of our training set. Subsequent exclusion of compounds with χ^0 greater than 9.5 from our training set and reapplication of their model resulted in a vastly improved fit (r^2=0.69). Although the compounds compiled by Suzuki *et al.* certainly constitute an eclectic set, one may argue that more diversity is needed especially regarding the drug-like molecules. Those included are relatively small compared to other drugs, and their presence results in the low range of χ^0 in their training set. Fortunately, our training set consists of larger drug molecules with much greater χ^0 values. In addition, Suzuki *et al.* also did not define the types of compounds that are appropriate for evaluation by their model. Since their training set is shown to be limited, a guideline for what drugs can and cannot be evaluated is needed.

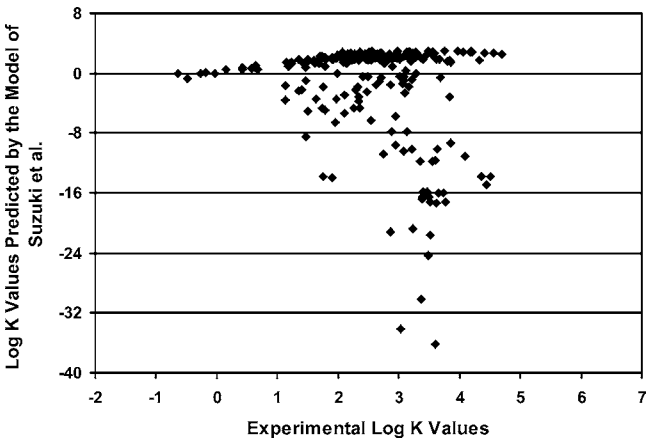

Fig. 5. Plot of Log K values determined by the model from Suzuki *et al.* (11) *versus* the experimental Log K values for all compounds of the training set.

The composition of the training set is also a concern for two models developed by Klein *et al.* (12,13). One is composed of first-order terms while the other contains both linear and non-linear terms. The two are based on the same training set of 70 compounds which consists exclusively of drug-like molecules. The set is larger than that of Suzuki *et al.*, but the diversity is reduced which limits application of the model to drug-like molecules. Unfortunately, the types of drugs that are suitable for evaluation by their models was not mentioned. Regardless, even if the model is used to evaluate such compounds, the possibility exists that some molecules could possess descriptor values that lie at the limits of the ranges of those values of their training set compounds. To reduce the possibility of such a situation, our training set was constructed with as many diverse compounds as possible.

Two models, one by Suzuki *et al.* (14) and the other from Katritzky *et al.* (15), were derived from the same robust training set of 218 diverse ligands. Both utilize a fragment based approach for determining the binding free energy between a compound and β-CD. In this technique, a molecule is analyzed for the presence of certain pre-defined fragments or functional groups. Each group has a specific contribution to the overall value of the binding free energy which is obtained by summing those individual contributions. The concern with these models lies not with their training set, but rather with the fragment based approach itself. This technique limits the types of compounds that can be evaluated. A molecule that contains very little or none of the fragments in the model training set cannot be properly analyzed. Unfortunately, neither study explicitly mentioned whether a compound with rare fragments could be confidently evaluated not did they provide guidelines on how to determine if a compound has an adequate number of represented fragments for evaluation. However, Suzuki *et al.* did provide a table of the frequency of each fragment in the training set. One can judge form this whether a drug is well-represented by the training set. Since we did not use the fragment based approach, this concern is not an issue for our models. Nevertheless, the two fragment based models have very good predictive abilities and can be confidently applied to compounds that have an adequate number of represented fragments.

CONCLUSIONS

In this study, two empirical models were developed to predict the binding between a ligand and β-CD. The binding

constant can be used to decide whether this CD is a suitable excipient for increasing the solubility of a drug candidate. The training set of the two models is larger and more diverse than those of previous models in the literature. Thus, the models developed here can be used with more confidence because it is more likely that any ligand examined will be similar to those in our training set. Our first model was developed with partial least squares regression. It contains 25 descriptors and has an r^2_{cv} of 0.79. Compounds whose descriptor values fall within one standard deviation of the average descriptor values of the training set (Table I) can be evaluated with it. Multiple linear regression was used to construct the second model. It contains four variables and has an r^2_{cv} of 0.65. The second model is recommended over the first primarily because less descriptor values are needed for calculation: Compounds whose SP, ALOGP, PHI, and *J* descriptor values fall within one standard deviation of the corresponding average descriptor values of the training set (Table I) can be examined with it. The calculated log *P* (ALOGP) was found to be the most important descriptor in both models. Removing it from either lowers their r^2_{cv} values to a greater extent than when any other descriptor is removed. This indicates that lipophilicity is the main property which governs a ligand's complexation with β-CD. Greater ALOGP values correlate with greater binding.

For future work, we seek to improve the models by increasing the size and diversity of the training set. This would increase the accuracy of the predictions and would also allow the models to be applied to a wider variety of drugs because the likelihood of a drug having descriptor values that are outliers from those of the training set is reduced. In addition, the training set could be divided into separate classes of compounds so that models can be made for specific types of drug. Other future work includes developing new models or modifying the current ones to predict whether a β-CD–ligand complex will be insoluble, *i.e.* show B-type solubility behavior. In addition, we will explore the possibility of developing empirical models to predict the binding between ligands and other CDs, including the more soluble substituted CDs such as sulfobutyl ether CD and hydroxypropyl CD.

ACKNOWLEDGEMENTS

The authors express their gratitude to Roche Inc. for their financial support of this project.

REFERENCES

1. R. Challa, A. Ahuja, J. Ali, and R. K. Khar. Cyclodextrins in drug delivery: an updated review. *AAPS PharmSciTech.* **6**:E329–E357 (2005) doi:10.1208/pt060243.
2. T. Loftsson, and D. Duchene. Cyclodextrins and their pharmaceutical applications. *Int. J. Pharm.* **329**:1–11 (2007) doi:10.1016/j.ijpharm.2006.10.044.
3. K. Connors. The stability of cyclodextrin complexes in solution. *Chem. Rev.* **97**:1325–1357 (1997) doi:10.1021/cr960371r.
4. K. Uekama, F. Hirayama, and T. Irie. Cyclodextrin drug carrier systems. *Chem. Rev.* **98**:2045–2076 (1998) doi:10.1021/cr970025p.
5. V. J. Stella, and R. A. Rajewski. Cyclodextrins: their future in drug formulation and delivery. *Pharm. Res.* **14**:556–567 (1997) doi:10.1023/A:1012136608249.
6. J. Pitha, and J. Pitha. Amorphous water-soluble derivatives of cyclodextrins: nontoxic dissolution enhancing excipients. *J. Pharm. Sci.* **74**:987–990 (1985) doi:10.1002/jps.2600740916.
7. G. Piel, I. Delneuville, and L. Delattre. In J. Szejtli and L. Szente (eds.), Proc. Eighth Int. Symp. on Cyclodextrins, Kluwer Academic, Dordrecht, 1996, pp. 487–490.
8. J. Szejtli. Past, present, and future of cyclodextrin research. *Pure Appl. Chem.* **76**:1825–1845 (2004) doi:10.1351/pac200476101825.
9. I. Shehatta, A. H. Al-Marzouqi, B. Jobe, and A. Dowaidar. Enhancement of aqueous solubility of itraconazole by complexation with cyclodextrins using supercritical carbon dioxide. *Can. J. Chem.* **83**:1833–1838 (2005) doi:10.1139/v05-181.
10. T. Loftsson. Cyclodextrins and the biopharmaceutics classification system of drugs. *J. Inclusion Phenom. Macrocyclic Chem.* **44**:63–67 (2002) doi:10.1023/A:1023088423667.
11. T. Suzuki, M. Ishida, and W. M. F. Fabian. Classical QSAR and comparative molecular field analyses of the host–guest interaction of organic molecules with cyclodextrins. *J. Comput.-Aided Mol. Des.* **14**:669–678 (2000) doi:10.1023/A:1008103122313.
12. C. T. Klein, D. Polheim, H. Viernstein, and P. Wolschann. A method for predicting the free energies of complexation between β-cyclodextrin and guest molecules. *J. Incl. Phenom. Macrocycl. Chem.* **36**:409–423 (2000) doi:10.1023/A:1008063412529.
13. C. T. Klein, D. Polheim, H. Viernstein, and P. Wolschann. Predicting the free energies of complexation between cyclodextrins and guest molecules: linear *versus* nonlinear models. *Pharm. Res.* **17**:358–365 (2000) doi:10.1023/A:1007565409407.
14. T. Suzuki. A nonlinear group contribution method for predicting the free energies of inclusion complexation of organic molecules with α- and β-cyclodextrins. *J. Chem. Inf. Comput. Sci.* **41**:1266–1273 (2001) doi:10.1021/ci010295f.
15. A. R. Katritzky, D. C. Fara, H. Yang, M. Karelson, T. Suzuki, V. P. Solov'ev, and A. Varnek. Quantitative structure–property relationship modeling of β-cyclodextrin complexation free energies. *J. Chem. Inf. Comput. Sci.* **44**:529–541 (2004) doi:10.1021/ci034190j.
16. Talete srl, DRAGON for Windows (Software for Molecular Descriptor Calculations). Version 5.4-2006-http://www.talete.mi.it/.
17. R. Todeschini, and V. Consonni. *Handbook of molecular descriptors*. Wiley-VCH, Weinheim (Germany), 2000.
18. M. V. Rekharsky, R. N. Goldberg, F. P. Schwarz, Y. B. Tewari, P. D. Ross, Y. Yamashoji, and Y. Inoue. Thermodynamic and nuclear magnetic resonance study of the interactions of alpha- and beta-cyclodextrin with model substances: phenethylamine, ephedrines, and related substances. *J. Am. Chem. Soc.* **117**:8830–8840 (1995) doi:10.1021/ja00139a017.
19. A. K. Ghose, and G. M. Crippen. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure–activity relationships I. Partition coefficients as a measure of hydrophobicity. *J. Comput. Chem.* **7**:565–577 (1986) doi:10.1002/jcc.540070419.
20. P. Ertl, B. Rohde, and P. Selzer. 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.* **43**:3714–3717 (2000) doi:10.1021/jm000942e.
21. Y. Inoue, T. Hakushi, Y. Liu, L. Tong, B. Shen, and D. Jin. Thermodynamics of molecular recognition by cyclodextrins. 1. Calorimetric titration of inclusion complexation of naphthalenesulfonates with alpha-, beta-, and gamma-cyclodextrins: enthalpy–entropy compensation. *J. Am. Chem. Soc.* **115**:475–481 (1993) doi:10.1021/ja00055a017.
22. E. Junquera, V. G. Baonza, and E. Aicart. Energetics of the encapsulation of *o*-, *m*-, and *p*-hydroxybenzoic acids by β-cyclodextrin and its methylated and hydroxypropylated derivatives in aqueous solution. *Can. J. Chem.* **77**:348–355 (1999) doi:10.1139/cjc-77-3-348.
23. K. Rajendrakumar, T. Pralhad, and S. Madhusudan. Comparative study on co-ground products of rofecoxib with β-cyclodextrin and its sulfobutyl ether-7 derivative in solution and in the solid state. *J. Inclusion Phenom. Macrocyclic Chem.* **49**:259–266 (2004) doi:10.1023/B:JIPH.0000048314.41966.65.
24. K. Uekama, Y. Hieda, F. Hirayama, H. Arima, M. Sudoh, A. Yagi, and H. Terashima. Stabilizing and solubilizing effects of sulfobutyl ether β-cyclodextrin on prostaglandin e1 ana-

logue. *Pharm. Res.* **18**:1578–1585 (2001) doi:10.1023/A:1013034615464.
25. P. Wallimann, T. Marti, A. Furer, and F. Diederich. Steroids in molecular recognition. *Chem. Rev.* **97**:1567–1608 (1997) doi:10.1021/cr960373b.
26. T. X. Xiang, and B. D. Anderson. Inclusion complexes of purine nucleosides with cyclodextrins: II, Investigation of inclusion complex geometry and cavity microenvironment. *Int. J. Pharm.* **59**:45–55 (1990) doi:10.1016/0378-5173(90)90063-A.
27. R. Carpignano, M. Marzona, E. Cattaneo, and S. Quaranta. QSAR study of inclusion complexes of heterocyclic compounds with β-cyclodextrin. *Anal. Chim. Acta.* **348**:489–493 (1997) doi:10.1016/S0003-2670(97)00152-9.
28. L. Liu, and Q.-X. Guo. Wavelet neural network and its application to the inclusion of β-cyclodextrin with benzene derivatives. *J. Chem. Inf. Comput. Sci.* **39**:133–138 (1999) doi:10.1021/ci980097x.
29. L. Liu, and Q.-X. Guo. Novel prediction for the driving force and guest orientation in the complexation of α- and β-cyclodextrin with benzene derivatives. *J. Phys. Chem. B.* **103**:3461–3467 (1999) doi:10.1021/jp984545f.
30. Y. Matsui, T. Nishioka, and T. Fujita. Quantitative structure–reactivity analysis of the inclusion mechanism by cyclodextrins. *Top. Curr. Chem.* **128**:61–89 (1985).
31. J. H. Park, and T. H. Nah. Binding forces contributing to the complexation of organic molecules with β-cyclodextrin in aqueous solution. *J. Chem. Soc. Perkin Trans.* **2**:1359–1362 (1994) doi:10.1039/p29940001359.
32. P. Geladi, and B. R. Kowalski. Partial least-squares regression: a tutorial. *Anal. Chim. Acta.* **185**:1–17 (1986) doi:10.1016/0003-2670(86)80028-9.
33. R. Kramer. *Chemometric techniques for quantitative analysis*. Marcel Dekker, New York, 1998.
34. W. Chen, C.-E. Chang, and M. K. Gilson. Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. *Biophys. J.* **87**:3035–3049 (2004) doi:10.1529/biophysj.104.049494.
35. G. L. Bertrand, J. R. Faulkner, S. M. Han, and D. W. Armstrong. Substituent effects on the binding of phenols to cyclodextrins in aqueous solution. *J. Phys. Chem.* **93**:6863–6867 (1989) doi:10.1021/j100355a057.
36. V. Zia, R. A. Rajewski, and V. J. Stella. Effect of cyclodextrin charge on complexation of neutral and charged substrates: comparison of $(SBE)_{7M}$-β-CD to HP-β-CD. *Pharm. Res.* **18**:667–673 (2001) doi:10.1023/A:1011041628797.
37. T. Higuchi, and K. Connors. Phase-solubility techniques. *Adv. Anal. Chem. Instrum.* **4**:117–212 (1965).
38. A. Yasri, and D. Hartsough. Toward an optimal procedure for variable selection and QSAR model building. *J. Chem. Inf. Comput. Sci.* **41**:1218–1227 (2001) doi:10.1021/ci010291a.
39. A. Trapani, A. Lopedota, N. Denora, V. Laquintana, M. Franco, A. Latrofa, and G. Trapani. A rapid screening tool for estimating the potential of 2-hydroxypropyl-β-cyclodextrin complexation for solubilization purposes. *Int. J. Pharm.* **295**:163–175 (2005) doi:10.1016/j.ijpharm.2005.02.013.