International Journal of Pharmaceutics 418 (2011) 207–216

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

journal homepage: www.elsevier.com/locate/ijpharm

Development of machine learning models of β-cyclodextrin and sulfobutylether-β-cyclodextrin complexation free energies

Alexei Merzlikine[a,*], Yuriy A. Abramov[a,**], Stacy J. Kowsz[a,b,1], V. Hayden Thomas[a], Takashi Mano[c]

[a] Department of Pharmaceutical Sciences, Pfizer Inc., Groton, CT, USA
[b] Rochester Institute of Technology, Rochester, NY, USA
[c] Department of Pharmaceutical Sciences, Pfizer Ltd., Sandwich, Kent, UK

ARTICLE INFO

Article history:
Received 31 January 2011
Received in revised form 16 March 2011
Accepted 24 March 2011
Available online 8 April 2011

Keywords:
Sulfobutylether-β-cyclodextrin
β-Cyclodextrin
Complexation constant
Machine learning methods
Cubist
Random Forest

ABSTRACT

A new set of 142 experimentally determined complexation constants between sulfobutylether-β-cyclodextrin and diverse organic guest molecules, and 78 observations reported in literature, were used for the development of the QSPR models by the two machine learning regression methods – Cubist and Random Forest. Similar models were built for β-cyclodextrin using the 233-compound dataset available in the literature. These results demonstrate that the machine learning regression methods can successfully describe the complex formation between organic molecules and β-cyclodextrin or sulfobutylether-β-cyclodextrin. In particular, the root mean square errors for the test sets predictions by the best models are low, 1.9 and 2.7 kJ/mol, respectively. The developed QSPR models can be used to predict the solubilizing effect of cyclodextrins and to help prioritizing experimental work in drug discovery.

1. Introduction

Modern drug discovery methods tend to advance large and hydrophobic molecules, which are likely to suffer from limited solubility and low bioavailability (Lipinski, 2000). Pharmaceutical industry is therefore increasingly interested in the drug delivery systems that can mitigate these risks. Cyclodextrins have been successfully used for drug solubilization both in research environment and in clinical use. For example, several commercial products – VFEND IV, Geodon IM, Abilify IM, and Sporanox – include substituted cyclodextrins in their formulations. Cyclodextrins can also stabilize the drug molecule against light, heat, or oxidation, and mask unwanted physiological effects (Hedges, 1998). Finally, their physical and biopharmaceutical properties can be tailored with relative ease by adding the appropriate functional groups to the 'parent' molecule (Szejtli, 1998). In particular, hydroxypropyl-β-cyclodextrin (HP-β-CD) and sulfobutylether-β-cyclodextrin (SBE-β-CD) have high aqueous solubility, are non-toxic, and have low oral bioavailability (Luke et al., 2010; Stella and He, 2008). Additional details on the pharmaceutical applications of cyclodextrins can be found in references (Brewster and Loftsson, 2007; Carrier et al., 2007; Davis and Brewster, 2004; Loftsson and Brewster, 1996; Loftsson and Duchene, 2007; Luke et al., 2010; Uekama et al., 1998).

Cyclodextrins increase solubility of drug molecules due to the formation of soluble inclusion complexes. A molecule of β-cyclodextrin is a cyclic oligosaccharide composed of 7 α-D-glycopyranose units. It has a toroidal (barrel-like) structure with the two openings having different diameters and with the hydroxyl groups situated around those openings (Szejtli, 1998). These hydroxyl groups make the exterior of a cyclodextrin molecule hydrophilic (Fig. 1) and fairly soluble. Moreover, these groups can be substituted with other functional groups such as hydroxypropyl or sulfobutyl ether in order to fine-tune the properties of a cyclodextrin molecule. The cyclodextrin's interior has low polarity (Fig. 1) and favors interaction with the lipophilic molecules (Connors, 1997). Thus, a non-polar molecule of the appropriate size may enter the inner cavity of cyclodextrin, displacing the 'high-energy' water molecules contained there into the bulk solution, and form a stable complex. Such mechanism of interaction defines the hydrophobic effect as a major driving force for the complex formation. In addition, dispersive interactions and hydrogen bonding (with the hydroxyl groups of cyclodextrin) also contribute to the complex stability (Connors, 1997).

Most commonly cyclodextrin forms a 1:1 inclusion complex with a drug molecule. The solubility of the resulting complex is typically higher than that of the drug molecule, and the apparent

* Corresponding author. Tel.: +1 860 686 9014; fax: +1 860 715 4473.
** Corresponding author. Tel.: +1 860 715 6005; fax: +1 860 441 5423.
E-mail addresses: alexei.merzlikine@pfizer.com (A. Merzlikine), yuriy.a.abramov@pfizer.com (Y.A. Abramov).
[1] Present address: University of California, Santa Barbara, CA, USA.

0378-5173/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2011.03.065

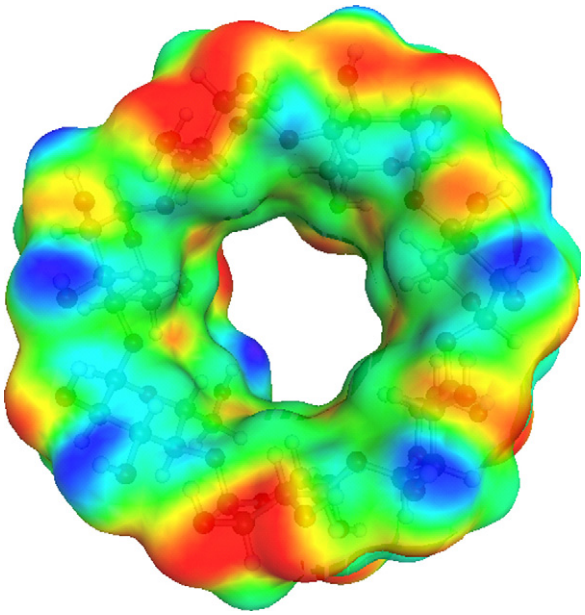

Fig. 1. Polarity of β-cyclodextrin molecule (top view) presented via its screening charge density in water medium. Calculation is done by COSMOtherm software (Eckert and Klamt, 2011) at the BP-SVP-AM1 COSMO level of theory. Red, blue and green colors represent acceptor, donor and non-polar (hydrophobic) surfaces, respectively.

drug solubility increases linearly with the cyclodextrin concentration. Hence, knowing the complexation constant allows one to predict the solubilizing effect offered by cyclodextrins. Inversely, the slope of the straight line in drug concentration/cyclodextrin concentration coordinates can be used to determine the complexation constant, as shown in the following equation:

$$K_{eq} = \frac{1}{S_0} \times \frac{\text{Slope}}{1 - \text{Slope}}, \quad (1)$$

where S_0 is the solubility of the guest molecule in the absence of cyclodextrin and Slope is the slope of the apparent solubility line (Connors, 1996; Higuchi and Connors, 1965). Such 1:1 complexes (commonly referred to as A_L-type) are relatively easy to analyze using phase-solubility method; however, the method fails for the complexes that deviate from 1:1 stoichiometry or whose solubility is less than that of the free drug (Connors, 1996, 1997; Higuchi and Connors, 1965).

Although many techniques exist for the experimental determination of drug–cyclodextrin complexation constant – the phase-solubility method (described above), spectroscopic methods (UV–vis, CD, NMR), capillary electrophoresis, and thermochemical methods – all of them require resources and compounds that may not be available in early stages of drug discovery. Pharmaceutical scientist frequently has to make decisions working with limited amounts of the Active Pharmaceutical Ingradient (API), or with virtual molecules, and under severe time constraints. This justifies a growing interest in developing the *in silico* methods for fast estimation of the cyclodextrin complexation energy with a drug molecule.

Applications of computational chemistry to studies of cyclodextrins have been reviewed by Lipkowitz (1998). Multiple *in silico* methods that are typically used to describe the cyclodextrin–guest molecule interactions were covered in the review: quantum mechanics (QM), molecular mechanics (MM), molecular dynamics (MD), Monte Carlo (MC), and quantitative structure–property (or activity) relationship (QSPR or QSAR). Each of these methods has its own advantages and disadvantages; however, only QSPR approach seems to provide a compromise between the speed and accuracy of the predictions. Thus, QSPR models can become a useful tool to help prioritizing the experimental work in an early drug discovery setting.

Recently, several QSPR methods were published describing complexation of β-cyclodextrin with a diverse set of organic molecules using different algorithms and descriptor sets (Katritzky et al., 2004; Perez-Garrido et al., 2009; Prakasvudhisarn et al., 2009; Suzuki, 2001; Suzuki et al., 2000). The modeling algorithms used in these studies included Comparative Molecular Field Analysis (CoMFA) (Suzuki et al., 2000), Group Contribution Method (GCM) (Suzuki, 2001), Multilinear Regression (MLR) (Katritzky et al., 2004; Perez-Garrido et al., 2009), Substructural Molecular Fragments (SMF) (Katritzky et al., 2004), Particle Swarm Optimization (PSO) and Support Vector Machines (SVMs) (Prakasvudhisarn et al., 2009).

A number of QSPR models for drug–β-cyclodextrin complexes have been published; however, the solubilization models for SBE-β-CD are not available. In this paper, we present a new dataset describing the equilibrium constants for 1:1 complexation between drug-like or simpler organic molecules and SBE-β-CD. We then report the development of the machine learning models based on this dataset using Cubist (Quinlan, 1993; Rulequest, 2010) and Random Forest (Breiman, 2001). Finally, we confirmed the applicability of these algorithms to β-cyclodextrin/small organic molecules complexation using literature data.

2. Materials and methods

2.1. Building the dataset

The dataset of the complexation constants between SBE-β-CD and a diverse set of organic molecules was created by combining the published data with the experimentally measured values. The literature data (complex formation free energies ΔG, equilibrium constants K_{eq}, or the parameters of the phase–solubility plots) were accepted if they belonged to 1:1 complexes. Free energies of complexation were used as is; the equilibrium constants were converted to the free energies using $\Delta G = -RT \ln K_x$, where K_x is a dimensionless equilibrium constant as opposed to apparent solubility K_{eq} having the dimension of M^{-1}. For aqueous solutions, $K_x \cong K_{eq} \times 55.5$. The majority of the reported equilibrium constants were measured at 25 °C and 37 °C; when the temperatures were not reported or reported as "room temperature", they were assumed to be 25 °C.

The dataset for the complexation between β-CD and small organic molecules was adapted from Katritzky et al. (2004).

A number of equilibrium constants for drug-like molecules were measured via phase–solubility technique (Connors, 1996; Higuchi and Connors, 1965) using SBE-β-CD (Captisol) purchased from CyDex Pharmaceuticals Inc. and the commercially available guest molecules.

Briefly, the experimentally determined concentration of the dissolved compound at equilibrium with its solid phase at several cyclodextrin concentrations was plotted against cyclodextrin concentration. For 1:1 complexes with the solubility higher than the intrinsic solubility of the guest compound, such plot becomes a straight line with the slope and intercept (solubility in the absence of cyclodextrin) related to the equilibrium constant according to Eq. (1).

Two series of measurements were performed. In the series 1 (S1 in Table 2), the excess of the solid compound was added to the 1, 2, 5, 10, 15, and 30 wt% aqueous solutions of SBE-β-CD. The samples were then equilibrated at 25 °C for a minimum of 24 h. Next, they were centrifuged at 14,000 rpm using Eppendorf 5804R

Table 1
Equilibrium constants for the complexation between drug-like molecules and SBE-β-CD reported in literature. Unless specified otherwise, the experiments were performed at 25 °C. Unspecified experimental temperatures, or the values reported to as "room temperature", were assumed to be 25 °C.

Compound	K_{eq}, M^{-1}	Comments	References
17-α-Methyltestosterone	12,933	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
1-Naphthol	1720.0		Kranack et al. (1998)
2-Naphthol	1304.0		Kranack et al. (1998)
5-Phenyl dithiolethione	10,705	Nonlinear phase solubility plot; the linear segment was used for K_{eq} determination; 37 °C	Dollo et al. (1999)
6-O-Benzylguanine	994	NCS 637037	Zia et al. (2000)
Acetohexamide	540.6	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Amlodipine	578.5	LC, CE; average of 2 enantiomers	Owens et al. (1998)
Anetholetrithione	12,834	Nonlinear phase solubility plot; the linear segment used for K_{eq} determination; 37 °C	Dollo et al. (1999)
Antalarmin	128.75	pH 2, cation	Sanghvi et al. (2009)
Benzthiazide	919.4	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Bupivacaine	149.3	Unionized; 37 °C	Dollo et al. (1998)
Butylmethoxydibenzoylmethane	2166.6	Nonlinear phase solubility plot, K of 1:1 interaction reported	Simeoni et al. (2004)
Carbamazepine	1035.0		Smith et al. (2005)
Carmustin	84.0		Ma et al. (2000)
Chlorpromazine	73,100		Okimoto et al. (1999a)
Chlorpromazine ion	32,100		Okimoto et al. (1999a)
Cinnarizine	69,700		Okimoto et al. (1996)
Danazol	374.6	22 °C	Jain and Adeyeye (2001)
Digitoxin	29,168	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Digoxin	11,851	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Dimethyldithiolethione	764.0	37 °C	Dollo et al. (1999)
Dithiolethione	364.0	37 °C	Dollo et al. (1999)
DY9760	3040	pH 7.4; 37 °C; effect of the ionic strength of the solution on K_{eq} strongly suggests electrostatic contribution	Nagase et al. (2001)
Estradiol	73,799		Okimoto et al. (2004)
Etomidate	445.0		McIntosh et al. (2004)
Flavopiridol	991.0		Li et al. (1999)
Flavopiridol ion	421.0		Li et al. (1999)
Fluasterone	216,129	1.55×10^{-7} M aq sol	Zhao et al. (1999)
Flurbiprofen	7996	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Fosphenytoin	41		Narisawa and Stella (1998)
Furosemide	338.3	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Griseofulvin	420.5	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Ibuprofen	2500	Averaged racemate and single isomer; unionized	Nerurkar et al. (2005)
Indomethacin	1590		Okimoto et al. (2004)
Indomethacin (ionized)	312.0		Okimoto et al. (2004)
Ketoprofen	1296	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Lidocaine	25.5	37 °C	Dollo et al. (1998)
Lorazepam	323.0		Holvoet et al. (2005)
m-Cresol	28.0		Adami et al. (2007)
Melphalan	360.0		Ma et al. (2000)
Methylprednisolone	717	Averaged value	706.0 (Larsen et al., 2005) 727 (Zia et al., 1997) (Zia et al., 2000)
Midazolam	700.0		Loftsson et al. (2001)
Midazolam, open form	425.0		Loftsson et al. (2001)
Naproxen	6757	Averaged value	3600 (Okimoto et al., 1996) (Okimoto et al., 2004) 9913 (non-phase-solubility) (Zia et al., 2000)
Neutral red, pH 5	750.0		Zhang et al. (2002)
Neutral red, pH 7.5	2300		Zhang et al. (2002)
Nifedipine	244.6	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
NSC-639829	3855		Jain et al. (2001)
Oxazepam	422	37 °C	Franco et al. (2004)
Papaverine	870	Averaged value for unionized molecule	1000.0 (Okimoto et al., 1996) 885 (Sotthivirat et al., 2007) 725 (Zia et al., 2000)
Phenol	128.0		Kranack et al. (1998)
Phenylalanine	32.4	pH 7	Miyajima et al. (2004)
Phenytoin	1170.0	Averaged value	1073 (Narisawa and Stella, 1998) and 1267 (Savolainen et al., 1998)
Phenytoin anion	476	pH 11	Savolainen et al. (1998)
Piroxicam	631.5	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Polythiazide	588.5	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Prazosin	14,750	Averaged value	21,315 (Sotthivirat et al., 2007) 11,733 unionized (Zia et al., 2000), (Zia et al., 2001) 11,202 cation (Zia et al., 2001)

Table 1 (*Continued*)

Compound	K_{eq}, M^{-1}	Comments	References
Prednisolone	1926	Averaged value	1513 (Okimoto et al., 2004) 2680 (Larsen et al., 2005) 1821 (Zia et al., 1997), 1691 (Sotthivirat et al., 2007)
Propofol	3725	30 °C, compare with 3686 at 37 °C	Babu and Godiwala (2004)
Methyl 4-(2-((1R,2R,3R)-3-hydroxy-2-((S,E)-3-hydroxy-4-(3-(methoxymethyl)phenyl)but-1-enyl)-5-oxocyclopentyl)ethylthio)butanoate	468.0		Uekama et al. (2001)
Quercetin	4032	30 °C	Jullian et al. (2007)
Rofecoxib	132.0		Rajendrakumar et al. (2004)
Sulfadimethoxine	304		Ueda et al. (1998)
Sulfathiazole	576.1	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Tacrolimus	420.0		Arima et al. (2001)
Testosterone	24,670	Averaged value	25855.0 (Sotthivirat et al., 2007); 23,486 (Okimoto et al., 1999b)
Thalidomide	86		Kale et al. (2008)
Thiabendazole	443.0		Okimoto et al. (1996)
Tolbutamide	230.05	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Trichlormethiazide	7.45	2-Point phase solubility plot; assuming linearity	Ueda et al. (1998)
Tryptophan	43.7	pH 7	Miyajima et al. (2004)
Tyrosine	31.9	pH 7	Miyajima et al. (2004)
Valdecoxib	1422	37 °C	Rajendrakumar et al. (2005)
Valproic acid	192	37 °C	Trapani et al. (2004)
Vinpocetine	340.0		Ribeiro et al. (2003)
Warfarin	5542	Averaged value	10100.0 (Okimoto et al., 1996) 4463 (Zia et al., 2000) 2063 unionized (Zia et al., 2001)
Warfarin anion	130		Zia et al. (2001)
Ziprasidone	7892	7892 (undissociated ziprasidone mesilate)	Kim et al. (1998)

centrifuge. The supernatant was collected, diluted, and analyzed by HPLC. Some compounds demonstrated deviation from linearity at the highest cyclodextrin concentration; in such cases, the deviating datapoints were excluded from the analysis because they did not correspond to the purely 1:1 association.

In the series 2 (S2), the solubility was determined in phosphate buffers with pH of 7.4 containing 0, 2, 4, and 6 wt% of SBE-β-CD. The buffer contained 38 g of $Na_2HPO_4 \bullet 12H_2O$ and 3.8 g of $NaH_2PO_4 \bullet 2H_2O$ per 1 L. The samples were stirred at 20 °C for 24 h, centrifuged for 15 min at 15,000 rpm, and the supernatant was diluted and analyzed by HPLC.

For both series, the equilibrium constant of the complex formation K_{eq} was determined from the linear phase–solubility plots corresponding to 1:1 complexes. We used the solubility value measured experimentally in the absence of cyclodextrins; in the cases when it was below the limits of detection, the intercept of the phase–solubility plot or a literature value was used.

2.2. Computational

2.2.1. Data preparation

218 data points from Katritzky et al. (2004) and 220 experimental observations reported in this study were used for development of QSPR models of β-CD and SBE-β-CD complexation free energies, respectively. All molecules were presented in the SD file format (Dalby et al., 1992) and were titrated *in silico* to pH 7 using Pfizer in-house protocols. This procedure was necessary to standardize model building and the future applications. When the experimental observations at the different pH values were available for a molecule, only the data relevant to the ionization at pH = 7 were retained. After that, the β-CD and SBE-β-CD datasets were split into training (90%) and test (10%) sets using a maximum dissimilarity algorithm, which allowed selection of the representative subsets of the original datasets. In addition, 15 experimental observations from the validation set of Suzuki (Suzuki, 2001) were used as a second test for the β-CD complexation model. Since the phase–solubility method does not provide information on the positive complexation free energies (i.e. repulsion), we assumed that all non-complexing systems ($K_{eq} = 0$) involving SBE-β-CD had $\Delta G = 0$ as opposed to $+\infty$.

2.2.2. Modeling approaches

We compared the results from the two different advanced machine learning regression methods – Random Forest (RF) (Breiman, 2001) and Cubist (Quinlan, 1993; Rulequest, 2010). Both methods were demonstrated to be suitable to model the data covering a very broad chemistry space with possible nonlinear relationships (Gao et al., 2008; Gupta et al., 2010; Palmer et al., 2007; Svetnik et al., 2003). In addition, both methods utilize built-in tools for selection of the important descriptors and thus are quite robust to overfitting problem.

Cubist is a tool for generating rule-based QSPR models which can be defined as a pairwise linear modeling method, except that the rules may overlap. Each rule is a conjunction of conditions associated with a linear expression. Cubist can also construct multiple models (committees); each of those is made up of several rule-based models. Predictions made by the each member of a committee for a target value of a case are averaged to give the final prediction. The prediction accuracy of a rule-based model can be improved by combining it with the nearest-neighbor (or instance-based) model. The latter predicts the target value of a new case by finding the *n* most similar cases in the descriptor space in the training data and averaging their target values. The importance of the individual descriptors can be estimated from the frequency of their use in the final model.

Random Forest (RF) is an ensemble of n_{tree} unpruned decision trees created by using bootstrap samples of the training data and random subset of m_{try} variables to define the best split at each node (tree fork) (Breiman, 2001). The bootstrap sample used during tree growth is a random selection with replacement from the molecules in the training set. Model performance for each tree is internally assessed with the prediction error of the data left-out in the bootstrap procedure (out-of-bag data). The average of these results for all trees provides an in situ cross-validation (out-of-bag validation). The RF prediction of new data is made by averaging the individual predictions of all the trees in the forest. In addition, RF has a built-in tool to measure the importance of individual descriptors across the training set.

The number of trees in the Random Forest in this study was set to a sufficiently large number of n_{tree} = 1000. The number of different descriptors tried at each split, m_{try}, was set to a default value of one third of the whole descriptor set (Svetnik et al., 2003).

2.2.3. Descriptors

Both Cubist and Random Forest methods utilize intrinsic (built-in) selection of the important descriptors and are generally not sensitive to the presence of the irrelevant features. Therefore, a relatively large set of two-dimensional (2D) descriptors was used in this study including Pfizer modified Molecular Operating Environment 2D (MOE 2D) set (Chemical Computing Group Inc., 2009, MOE 2009.10, http://www.chemcomp.com), Moriguchi & Blake descriptors (Moriguchi et al., 1992) and a set of in-house SMARTS keys (Lee et al., 2007; Tu and Li, 2004). In addition, Erlangen 2D descriptors (Bauknecht et al., 1996) were also used for building SBE-β-CD complexation free energy models. The total number of descriptors was decreased by exclusion of zero-variance and highly correlated descriptors – in the cases where the Pearson pairwise correlation coefficient exceeded the value of 0.85, one descriptor of the pair was removed.

2.2.4. Model selection and comparison

The model performance was evaluated using the predictions made for the test sets. Three statistical measures were evaluated – the root mean square error (RMSE), the mean absolute error (MAE) and the squared correlation coefficient between the observed and predicted data points (R^2):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2},$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i^{obs} - y_i^{pred}|,$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2}{\sum_{i=1}^{n}(y_i^{obs} - y_i^{obs,mean})^2},$$

where n is the set size, y_i^{obs} and y_i^{pred} are the observed and predicted values for molecule i.

In addition, prediction reliability was estimated using similarity to the training set of the compounds and the number of nearest neighbors (defined by similarity threshold of 0.7). The similarity matrix used in this study is atom pair similarity (Carhart et al., 1985; Sheridan et al., 2004). Better prediction is expected for the compounds with a larger number of the nearest neighbors and the higher maximum similarity below 1. Otherwise, there is a significant extrapolation, and the predicted values contain high uncertainty.

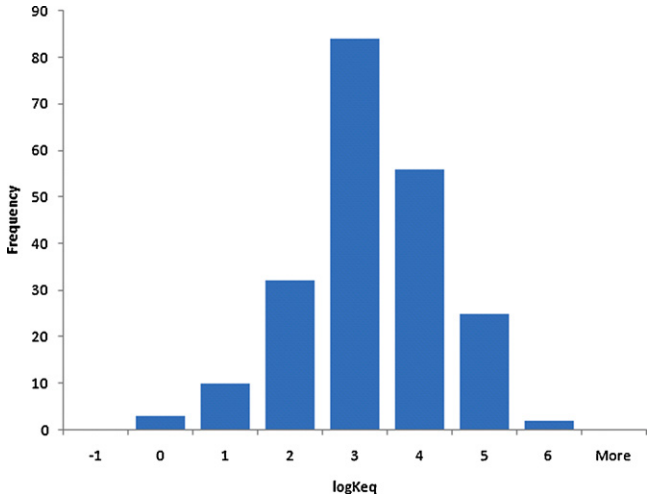

Fig. 2. Distribution of log K_{eq} of the dataset describing the complexation between the SBE-β-CD and the organic guest molecules reported in the current paper.

3. Results and discussion

The equilibrium constants for the complexation between SBE-β-CD and drug-like molecules reported in literature are summarized in Table 1. The measured constants are presented in Table 2.

Connors postulated that the logarithm of equilibrium constant for any non-covalent interaction is normally distributed over 5–6 units of magnitude with the standard deviation of 1 (Connors, 1997). Based on the properties of 721 complexes with β-CD, the average log K_{eq} was found to be 2.69, with the standard deviation of 0.89 (Connors, 1995) Our dataset describing 1:1 complexation of drug-like molecules with SBE-β-CD obeys this rule: log K_{eq} is distributed normally over the range of 5.7 units with the mean of 2.75 and the standard deviation of 1.05 (Fig. 2). This proves that the dataset reported in Tables 1 and 2 is representative of the drug–cyclodextrin interactions.

3.1. QSPR modeling results

The results of statistical performance of QSPR models for β-CD and SBE-β-CD complexation free energies are presented in Table 3. The scatter plots of the predicted vs. observed complexation free energies are presented in Figs. 3 and 4 for the best-performing models (RF or Cubist) only. Five most important descriptors for each model are listed in Table 4. As expected, many of these descriptors reflect the hydrophobic properties of the molecules.

The validation results show that the best models for β-CD and SBE-β-CD complexation free energies were obtained via Cubist and RF algorithms, respectively. The RF approach demonstrated more consistent performance by producing good models for both cyclodextrins, with Cubist failing to achieve a comparable result for SBE-β-CD model (Table 3). Overall, the performance of the best β-CD and SBE-β-CD models with the tests sets is quite good demonstrating the strength of the selected machine learning regression methods.

Generally, the predictive power of a QSPR model depends on the degree of similarity between the molecules to be predicted and the molecules in the training set, as measured by both the maximum similarity coefficient and the number of the nearest neighbors. Correspondingly, the largest errors in the β-CD and SBE-β-CD test sets prediction are made for the molecules having zero similarity to the training sets, as evidenced by griseofulvin, benzidine, and 3,5-dibromophenol in Figs. 3b, c and 4a, respectively. In spite of the fact that almost half of the compounds from the β-CD test sets have no

Table 2
Measured complexation constants between sulfobutylether-β-cyclodextrin and organic molecules. The measurements were taken at 25 °C (series 1, S1) and at 20 °C (series 2, S2). See experimental section for details.

Compound	K, M^{-1}	Comments
(−) Sulpiride	35	S2
1,2,3-Trichlorobenzene	31,567	S2
1-Naphthylamine	518	S2
1-Phenylpyrrole	555	S2
2-(1-Adamantyl)-4-methylphenol	148	S2; S_0 below limits of detection. K_{eq} calculated using the extrapolated value
2-(2-Aminophenyl)-benzothiazole	16,444	S2
2,2′,4,4′-Tetrahydroxybenzophenone	3006	S2
2-Naphthylamine	780	S2
2-Phenylquinoline	27,039	S2
3-(4-Methylbenzylidene)camphor	285	S2; S_0 below limits of detection. K_{eq} calculated using the extrapolated value
3,4-Dihydro-2(1H)-quinolinone	12	S2
3,5-Dibromophenol	18	S2; the compound is partially ionized under the experimental conditions
3,5-Dichlorophenol	9.2	S2; the compound is partially ionized under the experimental conditions
3-Benzoylpyridine	5104	S2
3-Tert-butylphenol	611	S2
4′-(Imidazol-1-yl)acetophenone	193	S2
4-(Trifluoromethoxy)phenylacetic acid	9.85	S2; the last point (6% SBE-β-CD) deviated from the linear trend and was excluded from analysis. The compound is partially ionized under the experimental conditions
4-(Trifluoromethyl)phenylacetic acid	20.5	S2; the compound is partially ionized under the experimental conditions
4,4′-Dihydroxybenzophenone	1718	S2; the last point (6% SBE-β-CD) deviated from the linear trend and was excluded from analysis
4,5-Diazafluoren-9-one	35	S2
4′-Cyclohexylacetophenone	21,277	S2
4′-Hydroxypropiophenone	632	S2; the compound is partially ionized under the experimental conditions
4-Tert-amylphenol	4516	S2
4-Tert-butylphenol	1886	S2
5-Aminosalicylic acid	2.23	S1
5-Fluorocytosine	0	S2
5-Fluorouracil	49	S2; the compound is partially ionized under the experimental conditions
6-Hydroxy-3,4-dihydro-2(1H)-quinolinone	183	S2
7-Hydroxy-3,4-dihydro-2(1H)-quinolinone	84	S2
Acetazolamide	103.97	S1
Acridine	2993	S2
Adenine	53.01	S1
Adenosine	17.24	S1
Atropine	5.3	S2; the compound is partially ionized under the experimental conditions
Azathioprine	115	S2
Bendroflumethiazide	548.5	S2; the compound is partially ionized under the experimental conditions
Benzamide	55.64	S1
Benzocaine	759.24	S1
Benzoic acid	497.63	S1
Betamethasone	213	S2
Biphenyl	951	S2
Bis-(4-hydroxyphenylmetnane)	5381	S2
Butamben	3522	S2
Butylparaben	2341	S2; the compound is partially ionized under the experimental conditions
Caffeine	0	S2
Carbazole	1935	S2
Chloramphenicol	234	Average of 144.9 (S1) and 323 (S2)
Chlorthalidone	310.92	S1
Chlorzoxazone	182.18	S1
Cimetidine	74	S2
Corticosterone	1303	S2
Cortisone	1124.22	S1
Cortisone-21-acetate	15,457	S2
Cytosine	0.46	S1
Dapsone	4356	S2
Deoxycorticosterone	11,456	S2
Dexamethasone	3253	S2
Diatrizoic acid	0.0	S1
Diazepam	496	S2
Dibenzofuran	1010	S2; S_0 below limits of detection. K_{eq} calculated using the extrapolated value.
Dibenzothiophene	1543	S2; S_0 below limits of detection. K_{eq} calculated using the extrapolated value.
Diuron	2069.02	S1
Equilin	39,489	S2
Estriol	25111.41	S1
Estrone	21755.33	S1
Ethylparaben	4433.54	S1
Ethynylestradiol	104,820	S2
Fenbufen	125	S2; the compound is partially ionized under the experimental conditions
Fluconazole	14	S2
Flufenamic acid	7032.34	S1
Fluocinolone acetonide	2514	S2
Folic acid	27.05	S1
Glafenine	6034	S2
Griseofulvin	105	S2
Guaifenesin	2.95	S1
Guanine	0.0	S1
Hydrochlorothiazide	2262.64	S1
Hydrocortisone	1247.34	S1; compare to literature value of 2516 (Zia et al., 2000)
Hydrocortisone-17-butyrate	2168	S2
Hydrocortisone-21-acetate	2600	S2
Hydroflumethiazide	44.31	S1
Ibuprofen, ionized	793	S1; buffered pH 6; pKa 4.4
Indapamide	368.28	S1
Indoprofen	1341.59	S1
Ketoprofen, ionized	87.53	S1
Khellin	12	S2
Meclofenamic acid	670	S2; the compound is partially ionized under the experimental conditions
Mefenamic acid	2228.56	S1
Menadione	290	S2

Table 2 (*Continued*)

Compound	K, M^{-1}	Comments
Methocarbamol	36.73	S1
Methyl-3-nitrobenzoate	135	S2
Methylparaben	3363	S2
Minoxidil	522	S2
Naphtalene	5493	S2
Naproxen, ionized	314	Average of 238 (Zia et al., 2001); 273 (S1); 432 (Okimoto et al., 2004)
Nitrofurantoin	18.06	S1
N-phenylanthranilic acid	14.5	S2; the compound is partially ionized under the experimental conditions
Omeprazole	93	S2
Phenacetin	285	S2
Phenanthrene	25,008	S2
Phenanthridine	2192	S2
Phenazine	894	S2
Phenolphthalein	6447	S1
Phenothiazine	2856	S2
Phenoxazine	577	S2
Phenylbutazone	0	S2
Pranlukast hemihydrate	205	S2; S_0 below limits of detection. K_{eq} calculated using the extrapolated value.
Praziquantel	893	S2
Prednisolone-21-acetate	621	S2
Primidone	314.6	S1
Progesterone	31,846	Average of 37048.35 (S1) and 26,644 determined by non-phase solubility method (Zia et al., 2000)
Propylparaben	1427	S2
Pyrazinamide	6.56	S1
Pyroquilon	244	S2
Quinidine	490	S2; the compound is partially ionized under the experimental conditions
Salicylamide	535.33	S1
Salicylic acid	17.08	S1
Spironolactone	20,408	Average of 15,816 (Jarho et al., 2000) 25,000 (S1, calculated using the intrinsic solubility from (Jarho et al., 2000))
Sulfacetamide	60.84	S1
Sulfadiazine	230.32	S1
Sulfamerazine	124.68	S1
Sulfamethazine	148.65	S1
Sulfamethoxazole	347.19	S1
Sulfisoxazole	2.4	S2
Tenoxicam	9	S2
Tetraethylthiuram disulfide	26,618	S1
Theobromine	1	S2
Theophylline	3.21	S1
Thiamphenicol	296.1	S1
Thianaphthene	8120	S2
Thianthrene	21,391	S2; S_0 below limits of detection. Literature value of 1.1×10^{-6} M was used to calculate K_{eq} (Stovall et al., 2005)
Thymine	0.7	S1
Tolnaftate	1844	S2; S_0 below limits of detection; the extrapolated solubility was negative. Literature value of 1.6×10^{-6} M was used to calculate K_{eq} (Peri et al., 1994)

Table 2 (*Continued*)

Compound	K, M^{-1}	Comments
Triamcinolone	703	S2
Triamcinolone acetonide	2904	S2
Triamterene	33	S2
Trimethoprim	87	S2
Uracil	0	S2
Uric acid	0.0	S1
Vanillin acetate	107	S2
Xanthene	8008	S2; S_0 below limits of detection; the extrapolated solubility was negative. Literature value of 6.2×10^{-6} M was used to calculate K_{eq} (Stovall et al., 2005)
Xanthine	0.0	S1

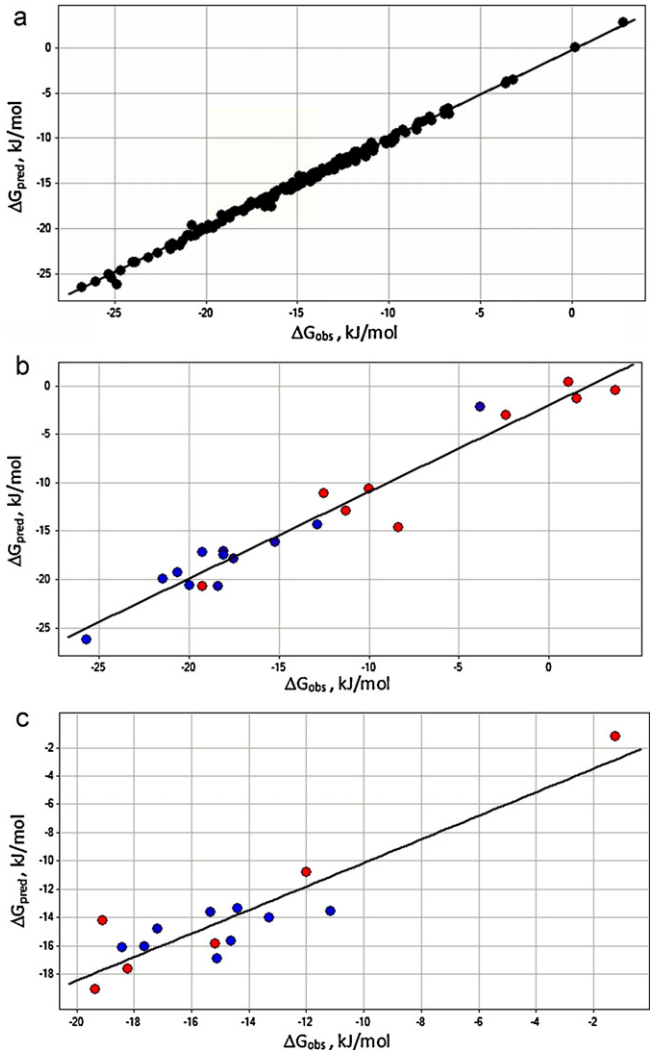

Fig. 3. Scatter plot of predicted (Cubist model) vs. observed β-CD complexation free energies for training (a), test 1 (b) and test 2 (c) sets. The color of the datapoints of the test set characterizes the highest similarity of the test molecules to those in the training set, ranging from red (zero similarity) to dark blue (highest similarity below 1).

Table 3
Statistical performance of the QSPR models for β-CD complexation free energies. The best model results according to the tests performance are highlighted in bold.

System	Statistical parameter	RF			Cubist		
		Training set	Test set 1	Test set 2	Training set	Test set 1	Test set 2
β-CD	R^2	0.978	0.912	0.816	0.996	**0.945**	**0.831**
	RMSE, kJ/mol	0.85	3.65	2.28	0.29	**2.09**	**1.92**
	MAE, kJ/mol	0.64	2.55	1.83	0.21	**1.54**	**1.52**
SBE-β-CD	R^2	0.938	**0.924**	–	0.704	0.912	–
	RMSE, kJ/mol	2.37	**2.71**	–	4.26	2.82	–
	MAE, kJ/mol	1.72	**2.12**	–	3.01	2.12	–

Table 4
Five most important descriptors in each best model.

	Symbol	Importance,[a] %	Description
β-CD (Cubist)	log *S*	69	Log of the aqueous solubility (mol/L).
	Q_VSA_HYD	17	Total hydrophobic van der Waals surface area.
	kS_sOH	12	Kier E-state for –OH
	BCUT_SLOGP_3	7	The BCUT descriptors using atomic contribution to log *P*
	PEOE_VSA_FPPOS	5	Fractional positive polar van der Waals surface area
SBE-β-CD (RF)	AC2D_A_QSIG_2	82	Autocorrelation vector with topological distance of 2 and atomic property of σ charge
	AC2D_A_QTOT_1	96	Autocorrelation vector with topological distance of 1 and atomic property of total charge
	AC2D_A_QPI_1	78	Autocorrelation vector with topological distance of 1 and atomic property of π charge
	log *P* (o/w)	66	Log of the octanol/water partition coefficient – a linear atom type model
	Q_VSA_FHYD	64	Fractional hydrophobic van der Waals surface area

[a] For the Cubist model the descriptor importance is presented via percentage of cases in the training data for which the descriptor appears in a condition of an applicable rule.

nearest neighbors in the training set (Fig. 3b and c), the overall prediction statistics of this model is very good (Table 3 and Fig. 3).

The best model for β-CD (Cubist) shows noticeably better training and validation results than the best model for SBE-β-CD (RF). Such behavior can be attributed to the more complicated nature of the SBE-β-CD/guest system. Most importantly, in contrast with the β-CD models derived using simpler organic molecules, SBE-β-CD model is developed for the larger and more complex (drug like) molecules. Additionally, SBE groups are negatively charged and might display electrostatic interaction with positively charged ligands outside the hydrophobic cavity thus introducing an alternative mechanism of complexation. The highest errors in the SBE-β-CD training set prediction (Fig. 4a) are obtained for the molecules with the observed zero K_{eq} values, especially for phenylbutazon and diatrizoic acid. As mentioned in the experimental section, the molecules with the $K_{eq} = 0$ were assigned zero free energy of complexation instead of infinity to enable their use in the model. The validation results suggest that this assumption could have been oversimplified, and such systems warrant further attention.

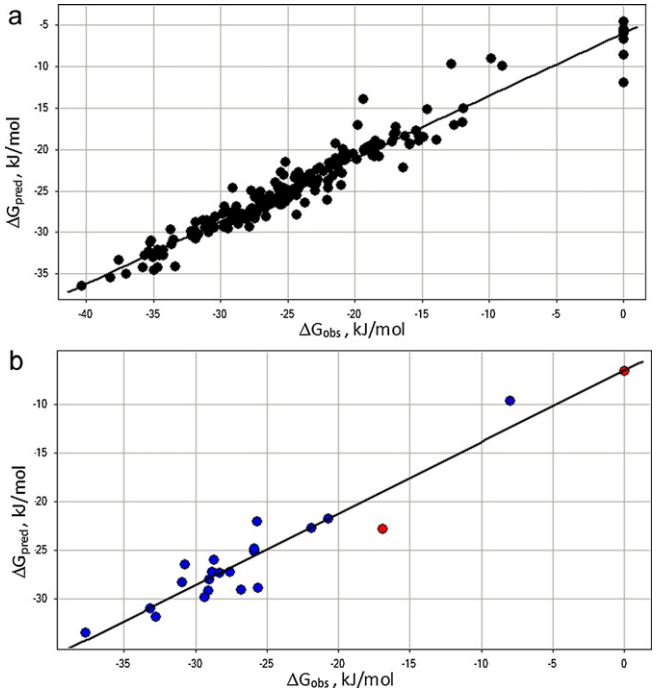

Fig. 4. Scatter plot of predicted (RF model) vs. observed SBE-β-CD complexation free energies for training (a) and test (b) sets. The color of the datapoints of the test set characterizes the highest similarity of the test molecules to those in the training set, ranging from red (zero similarity) to dark blue (highest similarity below 1).

4. Conclusions

In this study, we presented a new set of experimental data for SBE-β-CD complexation constants, which in combination with literature available observations were used for the development of the QSPR models by two machine learning regression methods – Cubist and Random Forest. Similar modeling was performed for β-CD complexation free energies based on the data available from the literature (Katritzky et al., 2004; Suzuki, 2001). The results of the modeling demonstrated successful applicability of the machine learning regression methods towards building β-CD and SBE-β-CD complexation models. Due to the relatively small size of the available data, the most reliable application of these models is limited to the certain chemical space. More experimental observations will allow expanding this space in the future studies.

Acknowledgments

The authors are grateful to Zhaohui Lei and Michael Lovdahl (Pfizer Inc.) for their help with optimizing the analytical measurements and to Joël Vacus and Isabelle Menier (Drugabilis Inc.) for solubility measurements and the discussion of the protocol.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijpharm.2011.03.065.

References

Adami, R.C., David, F., Wood, J.A., 2007. Antimicrobial Preservatives to Achieve Multidose Formulation Using Beta-cyclodextrins for Liquid Dosage Forms. Pharmacia & Upjohn.

Arima, H., Yunomae, K., Miyake, K., Irie, T., Hirayama, F., Uekama, K., 2001. Comparative studies of the enhancing effects of cyclodextrins on the solubility and oral bioavailability of tacrolimus in rats. J. Pharm. Sci. 90, 690–701.

Babu, M.K., Godiwala, T.N., 2004. Toward the development of an injectable dosage form of propofol: preparation and evaluation of propofol-sulfobutyl ether 7-beta-cyclodextrin complex. Pharm. Dev. Technol. 9, 265–275.

Bauknecht, H., Zell, A., Bayer, H., Levi, P., Wagener, M., Sadowski, J., Gasteiger, J., 1996. Locating biologically active compounds in medium-sized heterogeneous datasets by topological autocorrelation vectors: dopamine and benzodiazepine agonists. J. Chem. Inf. Comput. Sci. 36, 1205–1213.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

Brewster, M.E., Loftsson, T., 2007. Cyclodextrins as pharmaceutical solubilizers. Adv. Drug Deliv. Rev. 59, 645–666.

Carhart, R., Smith, D., Venkataraghavan, R., 1985. Atom pairs as molecular features in structure–activity studies: definition and application. J. Chem. Inf. Comput. Sci. 25, 64–73.

Carrier, R.L., Miller, L.A., Ahmed, I., 2007. The utility of cyclodextrins for enhancing oral bioavailability. J. Control. Release 123, 78–99.

Connors, K.A., 1995. Population characteristics of cyclodextrin complex stabilities in aqueous solution. J. Pharm. Sci. 84, 843–848.

Connors, K.A., 1996. Measurement of cyclodextrin complex stability constants. Compr. Supramol. Chem. 3, 205–241.

Connors, K.A., 1997. The stability of cyclodextrin complexes in solution. Chem. Rev. 97, 1325–1358.

Dalby, A., Nourse, J., Hounshell, W., Gushurst, A., Grier, D., 1992. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. J. Chem. Inf. Comput. Sci. 32, 244–255.

Davis, M.E., Brewster, M.E., 2004. Cyclodextrin-based pharmaceutics: past, present and future. Nat. Rev. 3, 1023–1035.

Dollo, G., Le Corre, P., Chevanne, F., Le Verge, R., 1998. Complexation between local anaesthetics and beta-cyclodextrin derivatives. Relationship between stability constants and in vitro membrane permeability of bupivacaine and lidocaine from their complexes. S.T.P. Pharm. Sci. 8, 189–195.

Dollo, G., Le Corre, P., Chollet, M., Chevanne, F., Bertault, M., Burgot, J.L., Le Verge, R., 1999. Improvement in solubility and dissolution rate of 1,2-dithiole-3-thiones upon complexation with beta-cyclodextrin and its hydroxypropyl and sulfobutyl ether-7 derivatives. J. Pharm. Sci. 88, 889–895.

Eckert, F., Klamt, A., 2011. COSMOTherm, C2.1_0111 ed. COSMOLogic GmbH&CoKG, Leverkusen, Germany.

Franco, M., Montenegro, L., Lopedota, A., Trapani, G., Puglisi, G., Liso, G., 2004. Effect of some hydrophilic cyclodextrins on the solubility, dissolution rate and in vitro percutaneous penetration of oxazepam. J. Drug Del. Sci. Technol. 14, 63–68.

Gao, H., Yao, L., Mathieu, H.W., Zhang, Y., Maurer, T.S., Troutman, M.D., Scott, D.O., Ruggeri, R.B., Lin, J., 2008. In silico modeling of nonspecific binding to human liver microsomes. Drug Metab. Dispos. 36, 2130–2135.

Gupta, R.R., Gifford, E.M., Liston, T., Waller, C.L., Hohman, M., Bunin, B.A., Ekins, S., 2010. Using open source computational tools for predicting human metabolic stability and additional absorption, distribution, metabolism, excretion, and toxicity properties. Drug Metab. Dispos. 38, 2083–2090.

Hedges, A.R., 1998. Industrial applications of cyclodextrins. Chem. Rev. 98, 2035–2044.

Higuchi, T., Connors, K.A., 1965. Phase–solubility techniques. Adv. Anal. Chem. Instrum. 4, 117–212.

Holvoet, C., Heyden, Y.V., Plaizier-Vercammen, J., 2005. Inclusion complexation of diazepam with different cyclodextrins in formulations for parenteral use. Die Pharmazie 60, 598–603.

Jain, A.C., Adeyeye, M.C., 2001. Hygroscopicity, phase solubility and dissolution of various substituted sulfobutylether beta-cyclodextrins (SBE) and danazol–SBE inclusion complexes. Int. J. Pharm. 212, 177–186.

Jain, N., Yang, G., Tabibi, S.E., Yalkowsky, S.H., 2001. Solubilization of NSC-639829. Int. J. Pharm. 225, 41–47.

Jarho, P., Vander Velde, D., Stella, V.J., 2000. Cyclodextrin-catalyzed deacetylation of spironolactone is pH and cyclodextrin dependent. J. Pharm. Sci. 89, 241–249.

Jullian, C., Moyano, L., Yanez, C., Olea-Azar, C., 2007. Complexation of quercetin with three kinds of cyclodextrins: an antioxidant study. Spectrochim. Acta 67, 230–234.

Kale, R., Tayade, P., Saraf, M., Juvekar, A., 2008. Molecular encapsulation of thalidomide with sulfobutyl ether-7 beta-cyclodextrin for immediate release property: enhanced in vivo antitumor and antiangiogenesis efficacy in mice. Drug Dev. Ind. Pharm. 34, 149–156.

Katritzky, A.R., Fara, D.C., Yang, H., Karelson, M., Suzuki, T., Solov'ev, V.P., Varnek, A., 2004. Quantitative structure–property relationship modeling of beta-cyclodextrin complexation free energies. J. Chem. Inf. Comput. Sci. 44, 529–541.

Kim, Y., Oksanen, D.A., Massefski Jr., W., Blake, J.F., Duffy, E.M., Chrunyk, B., 1998. Inclusion complexation of ziprasidone mesylate with beta-cyclodextrin sulfobutyl ether. J. Pharm. Sci. 87, 1560–1567.

Kranack, A.R., Bowser, M.T., Britz-McKibbin, P., Chen, D.D., 1998. The effects of a mixture of charged and neutral additives on analyte migration behavior in capillary electrophoresis. Electrophoresis 19, 388–396.

Larsen, K.L., Aachmann, F.L., Wimmer, R., Stella, V.J., Kjolner, U.M., 2005. Phase solubility and structure of the inclusion complexes of prednisolone and 6 alpha-methyl prednisolone with various cyclodextrins. J. Pharm. Sci. 94, 507–515.

Lee, P.H., Cucurull-Sanchez, L., Lu, J., Du, Y.J., 2007. Development of in silico models for human liver microsomal stability. J. Comput. Aided Mol. Des. 21, 665–673.

Li, P., Tabibi, S.E., Yalkowsky, S.H., 1999. Solubilization of flavopiridol by pH control combined with cosolvents, surfactants, or complexants. J. Pharm. Sci. 88, 945–947.

Lipinski, C.A., 2000. Drug-like properties and the causes of poor solubility and poor permeability. J. Pharmacol. Toxicol. Methods 44, 235–249.

Lipkowitz, K.B., 1998. Applications of computational chemistry to the study of cyclodextrins. Chem. Rev. 98, 1829–1874.

Loftsson, T., Brewster, M.E., 1996. Pharmaceutical applications of cyclodextrins. 1. Drug solubilization and stabilization. J. Pharm. Sci. 85, 1017–1025.

Loftsson, T., Duchene, D., 2007. Cyclodextrins and their pharmaceutical applications. Int. J. Pharm. 329, 1–11.

Loftsson, T., Gudmundsdottir, H., Sigurjonsdottir, J.F., Sigurdsson, H.H., Sigfusson, S.D., Masson, M., Stefansson, E., 2001. Cyclodextrin solubilization of benzodiazepines: formulation of midazolam nasal spray. Int. J. Pharm. 212, 29–40.

Luke, D.R., Tomaszewski, K., Damle, B., Schlamm, H.T., 2010. Review of the basic and clinical pharmacology of sulfobutylether-beta-cyclodextrin (SBECD). J. Pharm. Sci. 99, 3291–3301.

Ma, D.Q., Rajewski, R.A., Vander Velde, D., Stella, V.J., 2000. Comparative effects of (SBE)7m-beta-CD and HP-beta-CD on the stability of two anti-neoplastic agents, melphalan and carmustine. J. Pharm. Sci. 89, 275–287.

McIntosh, M.P., Schwarting, N., Rajewski, R.A., 2004. In vitro and in vivo evaluation of a sulfobutyl ether beta-cyclodextrin enabled etomidate formulation. J. Pharm. Sci. 93, 2585–2594.

Miyajima, M., Ozeki, T., Stella, V.J., 2004. Binding constants for aromatic amino acids and their derivatives with solfobutyl ether beta-cyclodextrin determined using capillary electrophoresis. J. Drug Del. Sci. Technol. 14, 383–387.

Moriguchi, I., Hirono, S., Liu, Q., Nakagome, I., Matsushita, Y., 1992. Simple method of calculating octanol/water partitioning coefficient. Chem. Pharm. Bull. 40, 127–130.

Nagase, Y., Hirata, M., Wada, K., Arima, H., Hirayama, F., Irie, T., Kikuchi, M., Uekama, K., 2001. Improvement of some pharmaceutical properties of DY-9760e by sulfobutyl ether beta-cyclodextrin. Int. J. Pharm. 229, 163–172.

Narisawa, S., Stella, V.J., 1998. Increased shelf-life of fosphenytoin: solubilization of a degradant, phenytoin, through complexation with (SBE)7m-beta-CD. J. Pharm. Sci. 87, 926–930.

Nerurkar, J., Beach, J.W., Park, M.O., Jun, H.W., 2005. Solubility of (+/−)-ibuprofen and S (+)-ibuprofen in the presence of cosolvents and cyclodextrins. Pharm. Dev. Technol. 10, 413–421.

Okimoto, K., Ohike, A., Ibuki, R., Aoki, O., Ohnishi, N., Irie, T., Uekama, K., Rajewski, R.A., Stella, V.J., 1999a. Design and evaluation of an osmotic pump tablet (OPT) for chlorpromazine using (SBE)7m-beta-CD. Pharm. Res. 16, 549–554.

Okimoto, K., Rajewski, R.A., Stella, V.J., 1999b. Release of testosterone from an osmotic pump tablet utilizing (SBE)7m-beta-cyclodextrin as both a solubilizing and an osmotic pump agent. J. Control. Release 58, 29–38.

Okimoto, K., Rajewski, R.A., Uekama, K., Jona, J.A., Stella, V.J., 1996. The interaction of charged and uncharged drugs with neutral (HP-beta-CD) and anionically charged (SBE7-beta-CD) beta-cyclodextrins. Pharm. Res. 13, 256–264.

Okimoto, K., Tokunaga, Y., Ibuki, R., Irie, T., Uekama, K., Rajewski, R.A., Stella, V.J., 2004. Applicability of (SBE)7m-beta-CD in controlled-porosity osmotic pump tablets (OPTs). Int. J. Pharm. 286, 81–88.

Owens, P., Fell, A., Coleman, M., Berridge, J., 1998. Effect of charged and uncharged chiral additives on the resolution of amlodipine enantioners in liquid chromatography and capillary electrophoresis. J. Chromatogr. A 797, 187–195.

Palmer, D.S., O'Boyle, N.M., Glen, R.C., Mitchell, J.B., 2007. Random forest models to predict aqueous solubility. J. Chem. Inf. Model 47, 150–158.

Perez-Garrido, A., Morales Helguera, A., Cordeiro, M.N.D.S., Garrido Escudero, A., 2009. QSPR modelling with the topological substructural molecular design approach: β-cyclodextrin complexation. J. Pharm. Sci. 98, 4557–4576.

Peri, D., Wyandt, C.M., Cleary, R.W., Hikal, A.H., Jones, A.B., 1994. Inclusion complexes of tolnaftate with β-cyclodextrin and hydroxypropyl β-cyclodextrin. Drug Dev. Ind. Pharm. 20, 1401–1410.

Prakasvudhisarn, C., Wolschann, P., Lawtrakul, L., 2009. Predicting complexation thermodynamic parameters of beta-cyclodextrin with chiral guests by using swarm intelligence and support vector machines. Int. J. Mol. Sci. 10, 2107–2121.

Quinlan, J.R., 1993. Combining instance-based and model-based learning. In: Proceeding of the Tenth International Conference on Machine Learning , Morgan Kaufmann.

Rajendrakumar, K., Madhusudan, S., Pralhad, T., 2005. Cyclodextrin complexes of valdecoxib: properties and anti-inflammatory activity in rat. Eur. J. Pharm. Biopharm. 60, 39–46.

Rajendrakumar, K., Pralhad, T., Madhusudan, S., 2004. Comparative study on co-ground products of rofecoxib with beta-cyclodextrin and its sulfobutyl ether-7 derivative in solution and in the solid state. J. Incl. Phenom. Macrocycl. Chem. 49, 259–266.

Ribeiro, L., Loftsson, T., Ferreira, D., Veiga, F., 2003. Investigation and physicochemical characterization of vinpocetine-sulfobutyl ether beta-cyclodextrin binary and ternary complexes. Chem. Pharm. Bull. 51, 914–922.

Rulequest, 2010. Cubist. RuleQuest Research, St. Ives, Australia.

Sanghvi, R., Mogalian, E., Machatha, S.G., Narazaki, R., Karlage, K.L., Jain, P., Tabibi, S.E., Glaze, E., Myrdal, P.B., Yalkowsky, S.H., 2009. Preformulation and pharmacokinetic studies on antalarmin: a novel stress inhibitor. J. Pharm. Sci. 98, 205–214.

Savolainen, J., Jarvinen, K., Matilainen, L., Jarvinen, T., 1998. Improved dissolution and bioavailability of phenytoin by sulfobutylether-beta-cyclodextrin and hydroxypropyl-beta-cyclodextrin complexation. Int. J. Pharm. 165, 69–78.
Sheridan, R.P., Feuston, B.P., Maiorov, V.N., Kearsley, S.K., 2004. Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. J. Chem. Inf. Comput. Sci. 44, 1912–1928.
Simeoni, S., Scalia, S., Benson, H.A., 2004. Influence of cyclodextrins on in vitro human skin absorption of the sunscreen, butyl-methoxydibenzoylmethane. Int. J. Pharm. 280, 163–171.
Smith, J.S., Macrae, R.J., Snowden, M.J., 2005. Effect of SBE7-beta-cyclodextrin complexation on carbamazepine release from sustained release beads. Eur. J. Pharm. Biopharm. 60, 73–80.
Sotthivirat, S., Haslam, J.L., Stella, V.J., 2007. Evaluation of various properties of alternative salt forms of sulfobutylether-beta-cyclodextrin, (SBE)7M-beta-CD. Int. J. Pharm. 330, 73–81.
Stella, V.J., He, Q., 2008. Cyclodextrins. Toxicol. Pathol. 36, 30–42.
Stovall, D.M., Acree, J.W.E., Abraham, M.H., 2005. Solubility of 9-fluorenone, thianthrene and xanthene in organic solvents. Fluid Phase Equilib. 232, 113–121.
Suzuki, T., 2001. A nonlinear group contribution method for predicting the free energies of inclusion complexation of organic molecules with alpha- and beta-cyclodextrins. J. Chem. Inf. Comput. Sci. 41, 1266–1273.
Suzuki, T., Ishida, M., Fabian, W.M., 2000. 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.
Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. 43, 1947–1958.
Szejtli, J., 1998. Introduction and general overview of cyclodextrin chemistry. Chem. Rev. 98, 1743–1754.
Trapani, G., Cutrignelli, A., Latrofa, A., Franco, M., Serra, M., Pisu, M.G., Biggio, G., Liso, G., 2004. Valproic acid–hydrophilic cyclodextrin complexes and valproic acid-solid dispersions: evaluation of their potential pharmaceutical use. Drug Dev. Ind. Pharm. 30, 53–64.
Tu, M., Li, D., 2004. An in Silico Model to Predict P-gp Substrate. The World Pharmaceutical Congress. Cambridge Healthtech Institute, Needham, MA, Philadelphia, PA.
Ueda, H., Ou, D., Endo, T., Nagase, H., Tomono, K., Nagai, T., 1998. Evaluation of a sulfobutyl ether beta-cyclodextrin as a solubilizing/stabilizing agent for several drugs. Drug Dev. Ind. Pharm. 24, 863–867.
Uekama, K., Hieda, Y., Hirayama, F., Arima, H., Sudoh, M., Yagi, A., Terashima, H., 2001. Stabilizing and solubilizing effects of sulfobutyl ether beta-cyclodextrin on prostaglandin E1 analogue. Pharm. Res. 18, 1578–1585.
Uekama, K., Hirayama, F., Irie, T., 1998. Cyclodextrin drug carrier systems. Chem. Rev. 98, 2045–2076.
Zhang, G., Shuang, S., Dong, Z., Dong, C., Pan, J., 2002. Investigation of the inclusion behavior of neutral red with β-cyclodextrin, hydroxypropyl-β-cyclodextrin, and sulfobutylether-β-cyclodextrin. Anal. Chim. Acta 474, 189–195.
Zhao, L., Li, P., Yalkowsky, S.H., 1999. Solubilization of fluasterone. J. Pharm. Sci. 88, 967–969.
Zia, V., Rajewski, R.A., Bornancini, E.R., Luna, E.A., Stella, V.J., 1997. Effect of alkyl chain length and degree of substitution on the complexation of sulfoalkyl ether beta-cyclodextrins with steroids. J. Pharm. Sci. 86, 220–224.
Zia, V., Rajewski, R.A., Stella, V.J., 2000. Thermodynamics of binding of neutral molecules to sulfobutyl ether beta-cyclodextrins (SBE-beta-CDs): the effect of total degree of substitution. Pharm. Res. 17, 936–941.
Zia, V., Rajewski, R.A., Stella, V.J., 2001. Effect of cyclodextrin charge on complexation of neutral and charged substrates: comparison of (SBE)7M-beta-CD to HP-beta-CD. Pharm. Res. 18, 667–673.