J Incl Phenom Macrocycl Chem (2011) 71:195–206
DOI 10.1007/s10847-011-9927-4

ORIGINAL ARTICLE

An alignment independent 3D-QSAR study for predicting the stability constants of structurally diverse compounds with β-cyclodextrin

Jahan B. Ghasemi · M. Salahinejad · M. K. Rofouei

Received: 2 October 2010 / Accepted: 17 January 2011 / Published online: 5 February 2011

Abstract For the first time, to our opinion, an alignment free, three dimensional quantitative structure activity relationships (3D-QSAR) of stability constants of a large and heterogeneous variety of organic substances with β-cyclodextrin was reported. GRIND methodology, where descriptors are derived from GRID molecular interaction fields (MIF), was used. After variable selection via fractional factorial design (FFD), PLS analysis was carried out, and a highly descriptive and predictive model was obtained. The model satisfied a set of rigorous validation criteria and performed well in the prediction of an external test set. The proposed model is also checked for free from chance correlation, reliability and robustness by permutation testing called progressive **y** scrambling. The obtained models confirmed that size and shape of the molecules as well as hydrophobic interactions are the main parameters influencing the stabilities of diverse compounds—β-cyclodextrin inclusion complexes.

Keywords β-cyclodextrin · Stability constant · 3D QSAR · GRIND descriptors

Introduction

Cyclodextrins (CDs) are cyclic oligomers that are widely used in host–guest applications. The most common natural CDs are α-CD, β-CD and γ-CD, consisting of six, seven and eight (1-4)-linked glucose units, respectively [1]. The CDs all form doughnut-shaped molecules with their hydroxyl groups on the outside of the molecule and a relatively nonpolar hole down the middle. Because of their structural features, CDs tend to form noncovalent inclusion complexes, and these have been used to improve properties such as solubility, dissolution rate, chemical and physical stability and bioavailability of poorly water-soluble compounds [2]. Therefore the CDs are the most suitable host molecules for the recognition of hydrophobic guest molecules, such as drugs, dyes, detergents, pesticides and etc. in aqueous media in a wide range of applications in industrial, pharmaceutical, agricultural and other fields [3–6].

The stability (association) constant value of host–guest complexes is a useful index of the binding strength of the complex and is of great importance for the understanding and evaluation of the inclusion complex formation [7]. Many attempts have been made to rationalize the stability of noncovalent host–guest inclusion complexes in terms of the molecular shape and size of guests, intermolecular interactions, or solvation effects as driving forces for CDs complexation [8]. The major interactions that have been put forward to account for the stabilities of CD inclusion complexes in aqueous solution were: release of 'high-energy' water from the CD cavity, relief of conformational strain energy possessed by the uncomplexed CD, the hydrophobic interaction, electrostatic interactions, hydrogen bonding, induction forces and the London dispersion force [9, 10]. CDs complex formation usually results from different combinations of these forces. One approach used for addressing the contribution each of these forces makes toward complexation is to rely on quantitative-structure–activity relationships (QSAR) [10, 11].

J. B. Ghasemi (✉)
Department of Chemistry, Faculty of Sciences, K.N. Toosi University of Technology, Tehran, Iran
e-mail: jahan.ghasemi@gmail.com

M. Salahinejad · M. K. Rofouei
Faculty of Chemistry, Tarbiat Moalem University, Tehran, Iran

QSAR, as a methodology that allows cost savings by reducing the laboratory resources needed and the time required to create and investigate new compounds with better complexing profile, permits one to separate one type of interaction from another and to do so quantitatively. QSAR/QSPR methods including 2D-QSAR [12, 13], 3D-QSAR and CoMFA [14, 15] have been applied to elucidate the most important factors influencing the host guest interactions and to predict the thermodynamic stability of CDs inclusion complexes.

The GRIND, alignment independent, interpretable and efficient to compute descriptors derived from GRID molecular interaction fields, was proved relevant in diverse structure–activity relationship studies. The aim of the present study, in continuation to our recent efforts to model the host:guest chemistry using QSAR approach [16, 17], is to develop, for the first time to our knowledge, a valid and predictive 3D-QSAR model, able to correlate and predict the complex stability constants between a large and diverse set of neutral molecules and β-cyclodextrin (β-CD), with the applicability of GRID independent descriptors (GRIND).

Materials and methods

Data set

A large and heterogeneous variety of 233 substances was extracted from the work of Suzuki [18] and the experimental data to be predicted is the β-CD complex stability constants (K) in water at 298 K taken from references therein. The set of guest molecules is structurally diverse and includes a large number of classes of organic compounds: aromatic hydrocarbons, alcohols, phenols, ethers, aldehydes, ketones, acids, esters, nitriles, anilines, halogenated compounds, heterocycles, nitro, sulfur and steroids and barbitals compounds and β-cyclodextrin as host. Log-transformed (log K) was used in the following QSAR modeling. Table 1 displays a complete list of the chemicals along with the reported experimental data. A statistical subset selection was made using most descriptive compound (MDC) method to select a balanced and chemically diverse test set. The MDC criterion privileges a selection scheme that weights the compounds according to their population density [19]. The training set of 173 molecules was used to adjust the parameters of the models, and the rest of molecules were used to evaluate model prediction ability. The model was externally validated with a test set of compounds, which were not considered for QSAR model generation (test set). The progressive scrambling method (maximum: 8 bins, minimum: 2 bins and critical point: 0.85) was carried out for the evaluation of the sensitivity of the 3D-QSAR model to chance correlations [20, 21]. Progressive scrambling with 2-6 components (latent variables) produces three statistical data, the predictivity (q^2) of the model, the calculated cross-validated standard error (cSDEP) and the sensitivity to perturbation ($d_q^{2\prime}/dr_{yy}^{\prime 2}$).

Molecular optimization and descriptor calculation

The 3D structure of the guest molecules was constructed using the standard tools available in the SYBYL 7.3 molecular modeling package (Tripose Inc., St. Louis, USA) running on a Red Hat Linux workstation 4.7. Energy minimization performed using the Tripos force field with a distance dependent dielectric and the Powell conjugate gradient algorithm with a convergence criterion of 0.01 kcal/mol Å. Partial atomic charges were calculated using the Gasteiger-Hückel method. GRid-INdependent Descriptors (GRIND) was calculated automatically using the software Pentacle, version1.05 [22]. The Pentacle software uses alignment independent descriptors derived from GRID molecular interaction fields (MIF). In this study we generated MIFs for DRY, N1, O and TIP probes defined as follows: the DRY probe represents hydrophobic interactions, N1 (amide) and O (carbonyl) probes represent hydrogen bond donor and acceptor groups, respectively, and the TIP probe represents a shape-field. All molecular interaction fields were computed with the grid resolution of 0.5 Å with the smoothing window 0.8 Å. AMANDA algorithm were used for the extraction of nodes from the obtained MIF, the distance and relative position of nodes were described by MACC2.

Results and discussion

It is widely believed that 3D descriptors should provide better descriptions of the binding interactions in host–guest chemistry. However, most 3D methods suffer from two constraints. First, the correct conformation of a molecule must be used, which may not even be the lowest energy conformation, to compare structurally different compounds; second, the compounds must be properly aligned, a step that is time-consuming and may introduce user bias [23]. The Grid Independent (GRIND) descriptors [24] were developed with the aim to overcome the alignment problem and were therefore selected for this study. The GRIND working procedure involves three steps: (a) computing a set of molecular interaction fields (MIFs), (b) filtering the MIFs to extract the most relevant regions, and (c) encoding of geometrical relationships into GRIND by computing the product of the interaction energy for each pair of filtered points (nodes). When MIF are computed for the guest

Table 1 Names, experimental and calculated stability constant (logK) of guest molecules

	Chemical name	CAS no	Log K (Exp)	Log K (Pred)
M1	Carbon tetrachloride	56-23-5	2.2	2.12
M2	Chloroform	67-66-3	1.43	1.46
M3	Methanol	67-56-1	−0.49	−0.22
M4	Acetaldehyde	75-07-0	−0.27	0.08
M5	Acetonitrile	75-05-8	−0.64	0.08
M6[a]	Ethanol	64-17-5	−0.03	−0.68
M7	1,2-Ethanediol	107-21-1	−0.19	−0.07
M8[a]	Acetone	67-64-1	0.42	−0.47
M9	1-Propanol	71-23-8	0.57	1.26
M10[a]	2-Propanol	67-63-0	0.63	0.77
M11	1,3-Propanediol	504-63-2	0.67	0.35
M12	Tetrahydrofuran	109-99-9	1.47	1.97
M13	Cyclobutanol	2919-23-5	1.18	1.69
M14	1-Butanol	71-36-3	1.22	1.08
M15[a]	2-Butanol	78-92-2	1.19	1.42
M16[a]	2-Methyl-1-propanol	78-83-1	1.62	1.36
M17	2-Methyl-2-propanol	75-65-0	1.68	1.60
M18	1,4-Butanediol	110-63-4	0.64	0.83
M19[a]	Diethylamine	109-89-7	1.36	0.83
M20	Cyclopentanol	96-41-3	2.06	2.18
M21	1-Pentanol	71-41-0	1.8	1.82
M22	2-Pentanol	6032-29-7	1.49	1.27
M23[a]	3-Pentanol	584-02-1	1.35	1.93
M24	2-Methyl-1-butanol	137-32-6	2.08	1.67
M25	2-Methyl-2-butanol	75-85-4	1.91	2.24
M26	3-Methyl-1-butanol	123-51-3	2.25	1.81
M27	3-Methyl-2-butanol	598-75-4	1.92	2.06
M28[a]	2,2-Dimethyl-1-propanol	75-84-3	2.71	2.41
M29	1,5-Pentanediol	111-29-5	1.22	1.90
M30	1,4-Dibromobenzene	106-37-6	2.97	3.11
M31[a]	1,4-Diiodobenzene	624-38-4	3.17	3.38
M32[a]	3,5-Dibromophenol	626-41-5	2.56	3.10
M33[a]	3,5-Dichlorophenol	591-35-5	2.07	3.18
M34	1-Chloro-4-nitrobenzene	100-00-5	2.15	2.57
M35	Fluorobenzene	462-06-6	1.96	2.16
M36	Bromobenzene	108-86-1	2.5	2.39
M37	Iodobenzene	591-50-4	2.93	2.11
M38	3-Fluorophenol	372-20-3	1.7	2.23
M39	4-Fluorophenol	371-41-5	1.73	2.07
M40	3-Chlorophenol	108-43-0	2.28	2.78
M41	4-Chlorophenol	106-48-9	2.61	2.60
M42[a]	3-Bromophenol	591-20-8	2.51	2.77
M43	4-Bromophenol	106-41-2	2.65	2.89
M44	3-Iodophenol	626-02-8	2.93	2.99
M45	4-Iodophenol	540-38-5	2.98	2.99
M46	Nitrobenzene	98-95-3	2.04	2.16
M47	4-Nitrophenol	100-02-7	2.39	2.91
M48	Benzene	71-43-2	2.23	2.47
M49	Phenol	108-95-2	1.98	2.15

Table 1 continued

	Chemical name	CAS no	Log K (Exp)	Log K (Pred)
M50[a]	Hydroquinone	123-31-9	2.05	1.77
M51	4-Nitroaniline	100-01-6	2.48	2.10
M52	Aniline	62-53-3	1.6	2.30
M53[a]	Sulfanilamide	63-74-1	2.76	2.30
M54	Cyclohexanol	108-93-0	2.67	2.68
M55	1-Hexanol	111-27-3	2.33	2.16
M56[a]	2-Hexanol	626-93-7	1.98	1.76
M57	2-Methyl-2-pentanol	590-36-3	1.99	1.89
M58[a]	3-Methyl-3-pentanol	77-74-7	2.15	2.53
M59	4-Methyl-2-pentanol	108-11-2	2.04	1.95
M60[a]	3,3-Dimethyl-2-butanol	464-07-3	2.75	2.90
M61	1,6-Hexanediol	629-11-8	1.69	1.80
M62[a]	Benzonitrile	100-47-0	2.23	1.70
M63	Benzothiazole	95-16-9	2.38	2.14
M64	4-Nitrobenzoic acid	62-23-7	2.34	2.76
M65	Benzaldehyde	100-52-7	1.78	2.12
M66[a]	Benzoic acid	65-85-0	2.12	1.76
M67	4-Hydroxybenzaldehyde	123-08-0	1.75	2.26
M68	4-Hydroxybenzoic acid	99-96-7	2.2	1.74
M69[a]	Benzyl chloride	100-44-7	2.45	2.49
M70	Toluene	108-88-3	2.09	2.36
M71	Benzyl alcohol	100-51-6	1.71	1.52
M72	Anisole	100-66-3	2.32	2.04
M73	m-Cresol	108-39-4	1.98	2.50
M74	p-Cresol	106-44-5	2.4	2.75
M75[a]	4-Methoxyphenol	150-76-5	2.21	2.55
M76	3-Methoxyphenol	150-19-6	2.11	1.84
M77	4-Hydroxybenzyl alcohol	623-05-2	2.16	1.67
M78	Hydrochlorothiazide	58-93-5	1.76	2.32
M79	N-Methylaniline	100-61-8	2.12	2.19
M80	1-Butylimidazole	4316-42-1	2.19	2.54
M81	1-Heptanol	111-70-6	2.85	3.05
M82	Phenylacetylene	536-74-3	2.36	2.55
M83	Thianaphthene	95-15-8	3.23	2.69
M84[a]	4-Fluorophenyl acetate	405-51-6	2.11	2.59
M85	3-Fluorophenyl acetate	701-83-7	1.91	2.27
M86	4-Chlorophenyl acetate	1878-66-6	2.5	3.06
M87	3-Chlorophenyl acetate	13031-39-5	2.44	2.34
M88	4-Bromophenyl acetate	1878-68-8	2.68	3.34
M89	3-Bromophenyl acetate	35065-86-2	2.67	2.31
M90[a]	4-Iodophenyl acetate	33527-94-5	3	3.08
M91	3-Iodophenyl acetate	61-71-2	3.07	2.58
M92	4-Nitrophenyl acetate	830-03-5	2.13	2.12
M93	Acetophenone	98-86-2	2.27	2.34
M94[a]	Phenyl acetate	122-79-2	2.1	2.39
M95[a]	Methyl benzoate	93-58-3	2.5	1.82
M96	3-Hydroxyacetophenone	121-71-1	2.06	2.25
M97[a]	4-Hydroxyacetophenone	99-93-4	2.18	2.63
M98	Acetoanilide	103-84-4	2.2	1.86

Table 1 continued

	Chemical name	CAS no	Log K (Exp)	Log K (Pred)
M99	p-Xylene	106-42-3	2.38	2.72
M100	Ethylbenzene	100-41-4	2.59	2.44
M101[a]	Phenetole	103-73-1	2.49	2.09
M102	2-Phenylethanol	60-12-8	2.15	2.96
M103	3-Ethylphenol	620-17-7	2.6	2.58
M104	4-Ethylphenol	123-07-9	2.69	3.74
M105	4-Ethoxyphenol	622-62-8	2.33	3.19
M106	3-Ethoxyphenol	621-34-1	2.35	2.31
M107[a]	3,5-Dimethoxyphenol	500-99-2	2.34	2.28
M108	N-Ethylaniline	103-69-5	2.34	2.08
M109	N,N-Dimethylaniline	121-69-7	2.36	2.61
M110	Barbital	57-44-3	1.78	1.88
M111[a]	Cyclooctanol	696-71-9	3.3	3.44
M112	1-Octanol	111-87-5	3.17	3.32
M113[a]	2-Octanol	4128-31-8	3.13	2.71
M114[a]	Quinoline	91-22-5	2.12	2.17
M115	3-Cyanophenyl acetate	55682-11-6	1.49	2.47
M116[a]	4-Hydroxycinnamic acid	4501-31-9	2.83	2.52
M117	Ethyl benzoate	93-89-0	2.73	2.06
M118[a]	4′-Hydroxypropiophenone	70-70-2	2.63	3.27
M119	3′-Hydroxypropiophenone	13103-80-5	2.61	2.73
M120	p-Tolyl acetate	140-39-6	2.49	2.63
M121[a]	3-Methylphenyl acetate	122-46-3	2.21	2.60
M122	4-Methoxyphenyl acetate	54771-60-7	2.45	2.63
M123	4-Propylphenol	645-56-7	3.55	3.80
M124	3-Propylphenol	621-27-2	3.28	2.83
M125	4-Isopropylphenol	99-89-8	3.58	3.40
M126	3-Isopropylphenol	618-45-1	3.44	2.43
M127	4-Isopropoxyphenol	7495-77-4	2.86	2.86
M128	2-Norbornaneacetate	1007-01-8	3.59	2.74
M129	1-Benzylimidazole	4238-71-5	2.61	2.65
M130	m-Methylcinnamic acid	3029-79-6	2.93	2.77
M131	4-Ethylphenyl acetate	3245-23-6	2.83	3.10
M132	3-Ethylphenyl acetate	3056-60-8	2.68	2.54
M133	4-Ethoxyphenyl acetate	22545-15-9	2.54	2.63
M134[a]	3-Ethoxyphenyl acetate	52600-91-6	2.49	2.53
M135	Allobarbital	52-43-7	1.98	2.21
M136[a]	4-n-Butylphenol	1638-22-8	3.97	4.61
M137	3-n-Butylphenol	4074-43-5	3.76	3.50
M138	3-Isobutylphenol	30749-25-8	4.18	3.24
M139[a]	4-sec-Butylphenol	99-71-8	2.7	3.88
M140	3-sec-Butylphenol	3522-86-9	4.06	3.22
M141	4-tert-Butylphenol	98-54-4	4.56	3.64
M142	3-tert-Butylphenol	585-34-2	4.41	3.33
M143	Menadion	57-28-5	2.27	2.30
M144[a]	Sulfapyridine	144-83-2	2.48	2.38
M145	Sulfamonomethoxine	1220-83-3	2.48	2.53
M146	Sulfisoxazole	127-69-5	2.32	2.50
M147	4-n-Propylphenyl acetate	61824-46-2	3.15	3.23

Table 1 continued

	Chemical name	CAS no	Log K (Exp)	Log K (Pred)
M148[a]	3-n-Propylphenyl acetate	717918-70-2	3.28	3.09
M149[a]	4-Isopropylphenyl acetate	2664-32-6	2.88	2.80
M150	3-Isopropylphenyl acetate	4861-85-2	3.36	2.42
M151	4-n-Amylphenol	14938-35-3	4.19	4.09
M152	4-tert-Amylphenol	80-46-6	4.7	4.14
M153[a]	Carbutamide	339-43-5	2.29	2.14
M154	Pentobarbital	76-74-4	3.01	2.96
M155	Amobarbital	57-43-2	3.07	2.58
M156	Thiopental	76-75-5	3.28	2.80
M157	Dibenzofuran	132-64-9	2.97	2.97
M158	Dibenzothiophene	132-65-0	3.48	3.26
M159	Phenazine	92-82-0	2.41	2.57
M160	Thianthrene	92-85-3	3.57	3.10
M161	Carbazole	86-74-8	2.44	2.94
M162	Phenoxazine	135-67-1	2.69	2.86
M163	Phenothiazine	92-84-2	2.73	3.03
M164	Furosemide	54-31-9	1.78	2.24
M165	Phenobarbital	50-06-6	3.22	2.71
M166[a]	Sulfisomidine	515-64-0	2.1	1.68
M167	Sulfamethomidine	3772-76-7	2.33	2.60
M168	Sulfadimethoxine	122-11-2	2.26	2.50
M169	4-n-Butylphenyl acetate	14377-19-6	3.62	3.15
M170[a]	3-n-Butylphenyl acetate	122-43-0	3.66	3.40
M171	3-Isobutylphenyl acetate	66622-47-7	3.83	3.43
M172[a]	4-tert-Butylphenyl acetate	3056-64-2	2.71	3.06
M173	Cyclobarbital	52-31-3	2.71	3.06
M174	Hexobarbital	56-29-1	3.08	3.25
M175	1-Adamantaneacetate	4942-47-6	4.32	2.91
M176	Acridine	260-94-6	2.33	2.32
M177	Phenanthridine	229-87-8	2.57	2.18
M178[a]	Xanthene	92-83-1	3.1	2.30
M179	N-Phenylanthranilic acid	91-40-7	2.89	3.04
M180	Mephobarbital	115-38-8	3.16	2.20
M181	4-n-Amylphenyl acetate	17713-58-5	3.8	3.29
M182[a]	Flufenamic acid	530-78-9	2.35	2.80
M183	Meclofenamic acid	644-62-2	2.67	3.23
M184	Nitrazepam	146-22-5	1.97	2.94
M185	Flurbiprofen	5104-49-4	3.69	3.08
M186[a]	Sulfaphenazole	526-08-9	2.4	2.76
M187	Bendroflumethiazide	73-48-3	1.9	1.96
M188	Mefenamic acid	61-68-7	2.49	3.24
M189	Acetohexamide	968-81-0	2.94	3.52
M190	Fludiazepam	3900-31-0	2.33	2.32
M191	Nimetazepam	2011-67-8	1.73	2.80
M192	Fenbufen	36330-85-5	2.63	3.07
M193	Ketoprofen	22071-15-4	2.85	3.20
M194[a]	Medazepam	2898-12-6	2.53	2.43
M195[a]	Progabide	62666-20-0	3.83	2.54
M196	Griseofulvin	126-07-8	1.47	2.92

Table 1 continued

	Chemical name	CAS no	Log K (Exp)	Log K (Pred)
M197[a]	Tolnaftate	2398-96-1	3.35	3.03
M198	Prostacyclin	63859-31-4	2.94	3.35
M199	Triamcinolone	124-94-7	3.37	3.78
M200[a]	Cortisone	53-06-5	3.56	4.06
M201	Prednisolone	50-24-8	3.56	3.62
M202	Hydrocortisone	50-23-7	3.6	4.29
M203	Corticosterone	53-06-5	3.85	4.03
M204	Dexamethasone	50-02-2	3.65	3.83
M205[a]	Betamethasone	378-44-9	3.73	3.81
M206	Paramethasone	53-33-8	3.4	3.55
M207	Cortisone-21-acetate	50-04-4	3.62	3.38
M208[a]	Prednisolone-21-acetate	52-21-1	3.76	3.42
M209	Hydrocortisone-21-acetate	50-03-3	3.51	3.67
M210	Fluocinolone acetonide	67-73-2	3.48	3.77
M211[a]	Triamcinolone acetonide	76-25-5	3.51	3.78
M212	Spironolactone	52-01-7	4.44	3.62
M213	Dehydrocholic acid	81-23-2	3.38	4.07
M214	Chenodeoxycholic acid	474-25-9	4.36	3.78
M215	Ursodeoxycholic acid	128-13-2	4.51	3.39
M216[a]	Cholic acid	81-25-4	3.5	3.78
M217	Hydrocortisone-17-butyrate	94-25-7	3.23	2.97
M218	Cinnarizine	7002-58-6	3.64	2.80
M219	Cycloheptanol	502-41-0	3.23	2.78
M220	2-Methoxyethanol	109-86-4	0.22	0.14
M221	3-Hydroxycinnamic acid	588-30-7	2.56	1.60
M222[a]	Ethyl 4-hydroxybenzoate	120-47-8	3.01	2.32
M223	Ethyl 4-aminobenzoate	94-09-7	2.69	2.43
M224[a]	4-Methylcinnamic acid	1866-39-3	2.65	2.91
M225	Sulfadiazine	68-35-9	2.52	2.07
M226	L-a-O-benzylglycerol	56552-80-8	2.11	2.11
M227[a]	Sulfamerazine	127-79-7	1.97	2.58
M228	Butyl 4-hydroxybenzoate	94-26-8	3.39	2.90
M229[a]	Butyl 4-aminobenzoate	94-25-7	3.19	2.98
M230	Benzidine	92-87-5	3.35	3.49
M231[a]	Triflumizole	99387-89-0	2.66	3.12
M232	Diazepam	439-14-5	2.33	2.16
M233	Prostaglandine E2	363-24-6	3.09	3.30

[a] Test set

molecules, the region showing favorable energies of interaction represent positions where host molecules would interact favorably with guest molecules [25].

Model construction

Chemometric analysis was carried out using the statistical tools included in Pentacle software. The GRIND descriptors were related to stability constants by means of partial least square (PLS) analysis [26]. The GRIND descriptors were the **X** variables and stability constants, after log-transformation, were the **y** variable. The optimum number of PLS component (latent variables, LV) was chosen by monitoring changes in the model's predicting index (q^2_{loo}, leave one out) evaluated by applying the cross-validation procedure available in Pentacle. The PLS coefficients represent the contribution of each single descriptor to the model only with respect to **y** variable. Validation of the model was performed internally and externally using cross-validation method and test set respectively. In the final model, a total of 501 descriptors were derived after variable selection via fractional factorial design (FFD). The

PLS analysis resulted in a model with six latent variables and an $R^2_{train} = 0.87$ with a standard deviation of the error of calculation (SPEC) of 0.34. We tested the predictivity of the obtained model with a test set of 60 diverse guest molecules, for structure see Table 1. The cross validation of the model by LOO technique yielded q^2 values of 0.75 and by external validation of the model the R^2_{Pred} value of 0.74 was obtained. The experimental and calculated log K values are listed in Table 1.

Model validation and applicability domain

In order to check the reliability of the PLS model generated in our studies to chance correlation, we have used the progressive scrambling method [21]. In this approach, small random perturbations are introduced into training set and the statistical results, the perturbation prediction (q^2), the calculated cross-validated standard error of prediction (cSDEP) as the function of the correlation coefficient between the true values (**y**) of the dependent variables and the perturbed values (**y**′) of the dependent variables, and the slope of q^2 (cross validated correlation coefficient) with respect correlation of the original dependent variables against the perturbed dependent variables ($d_q^{2\prime}/dr_{yy}{}'^2$), are calculated. The obtained values of the sensitivity to the perturbation $d_q^{2\prime}/dr_{yy}{}'^2$, the prediction and cSDEP produced by a progressive scrambling analyses were 0.8779, 0.513 and 0.654 for the model respectively. These values confirm the robustness and independent of chance correlation of the model.

The analysis of the chemical applicability domain (AD) of the obtained model and the reliability of the predictions are also verified by the leverage approach, which is based in computing the leverage, h*, for each compound for which the QSPR model is used to predict the property under study [27, 28]. The warning leverage is generally fixed at 3 k/n, where k is the number of the model parameters plus 1 and n is the number of training set compounds. The analysis of the applicability domain of the model, Fig. 1, reveals the presence of just one chemical as outlier in the training set, namely 4-ethylphenyl acetate (M131). The statistics' of the model were not significantly affected when omitting this compound. It is also important to note that the validation compounds which were not used for model development are predicted with similar accuracy of the training compounds.

Interpretation of descriptors

Figure 2 shows the PLS coefficient plot indicated the most important pairs of nodes that contribute negatively or positively to the stability constants. A first inspection of the PLS coefficients plot enabled us to select some **X** variables

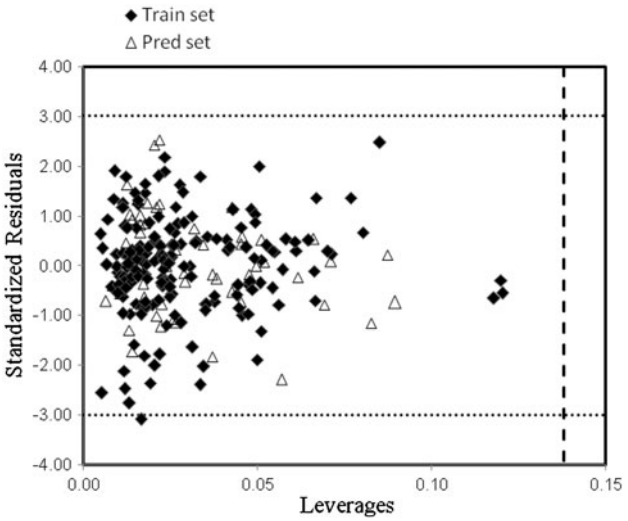

Fig. 1 Plot of standardized residuals versus leverages. *Dotted lines* represent ±3 standardized residual, *dash line* represents warning leverage (h* ≈ 0.14)

with the highest impact on the **y** variable. The largest peaks were related to the TIP probe (correlograms TIP-TIP, DRY-TIP and N1-TIP), which represent shape and size of the molecules. The next effective peaks were related to the DRY probe, suggesting that in a preliminary analysis the size and shape of the molecules as well as presence and orientation of hydrophobic groups were crucial for the stability constants of compounds here investigated with *β*-CD. To gain a deeper insight into the models, the variables with highest impact on **y** variable were inspected in more detail. Variable 188, TIP-TIP: distance 6.8–7.2 Å, explained the largest impact on stability constant with an inverse relationship. Small size molecules like: M3, M4, M5, and M7 (methanol, acetaldehyde, acetonitrile and 1,2 ethandiol) had the largest value of variable 188, but for a medium size molecule like M88 (4-bromophenyl acetate) has a low value of that variable. The distance node of variable 188 (6.8–7.2) is very near to inner diameter of *β*-CD (6.2–7.8 Å). In general, the efficiency and selectivity in host: guest binding strongly depends on shape and pre-organization within the host molecule, and the size-match of the host cavity to the guest [29].

The size-fit effect appears to play a subsidiary role in the inclusion complexation of the host: guest molecules. The fit of the entire or at least a part of the guest molecule in the cyclodextrin host cavity determines the stability of the inclusion complex and the selectivity of the complexation process [30]. Small compounds, smaller than cavity size of CD, easily resort in the cavity of the *β*-CD, they can't have good interaction with inner side of the *β*-CD cavity and therefore they have very low stability constants. Interestingly, the most variables of TIP-TIP correlogram that have negatively correlated bars are located on the left side of the

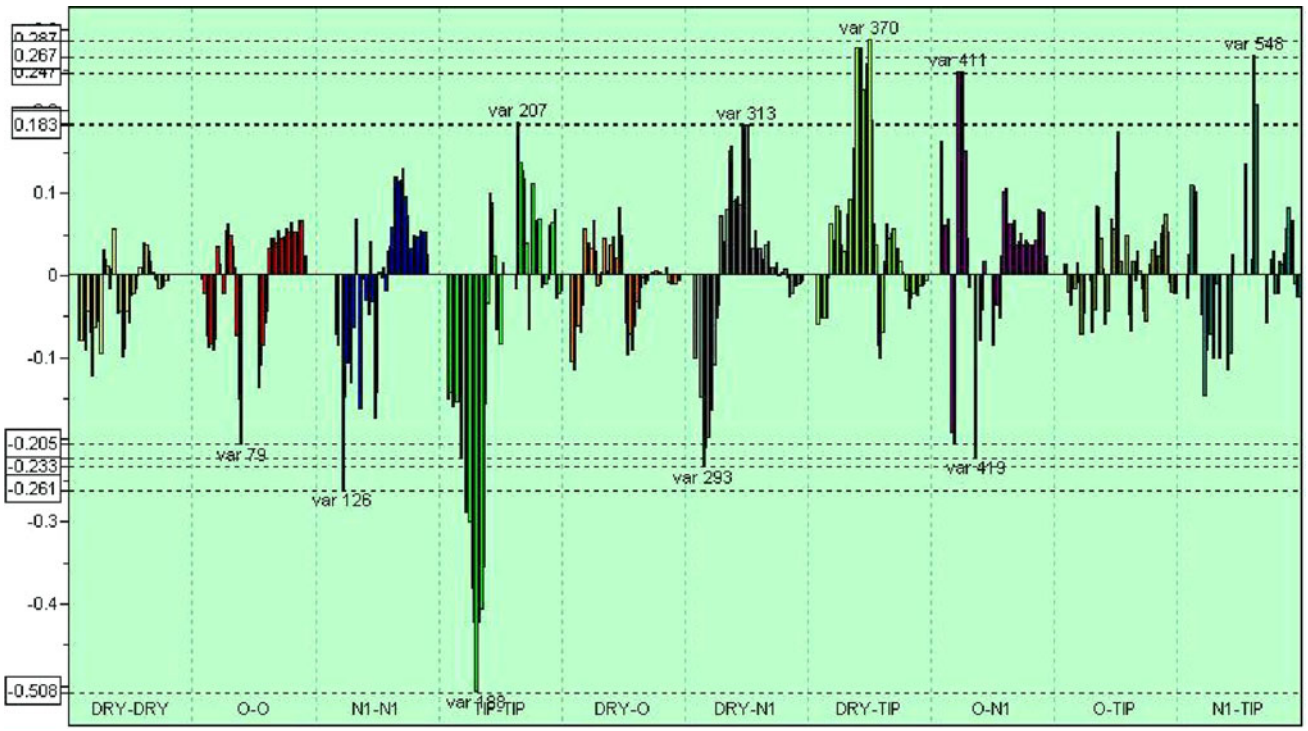

Fig. 2 PLS coefficient plot

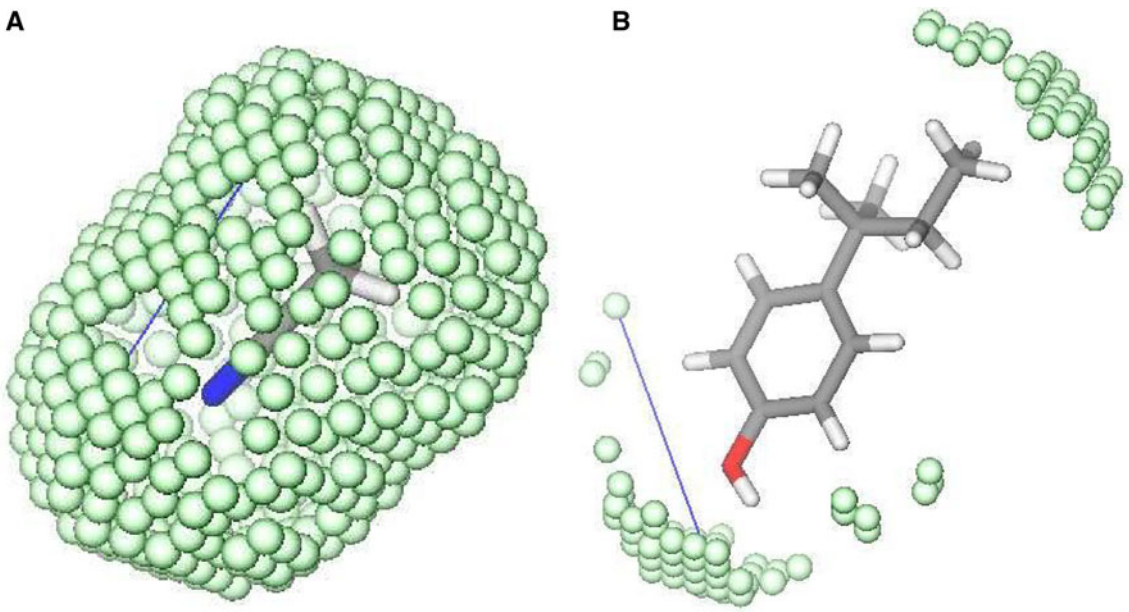

Fig. 3 Graphical display of GRIND variable 188 of the TIP-TIP auto-correlogram for (**a**): the most inactive (M5) and (**b**): the most active (M152) compound

correlogram and variables with positively impact on the **y** are positioned on larger node–node distance, i.e. on right side of correlogram.

Figure 3 shows graphical display of variable 188 for the most inactive (M5) and the most active (M152) compound. Variable 370 indicated a significant distance of 11.2–11.6A° between TIP and DRY nodes which has positive correlation with stability constants. It is well known that DRY probe favorably interact with different types of π systems (aromatic or vinyl type), but have not

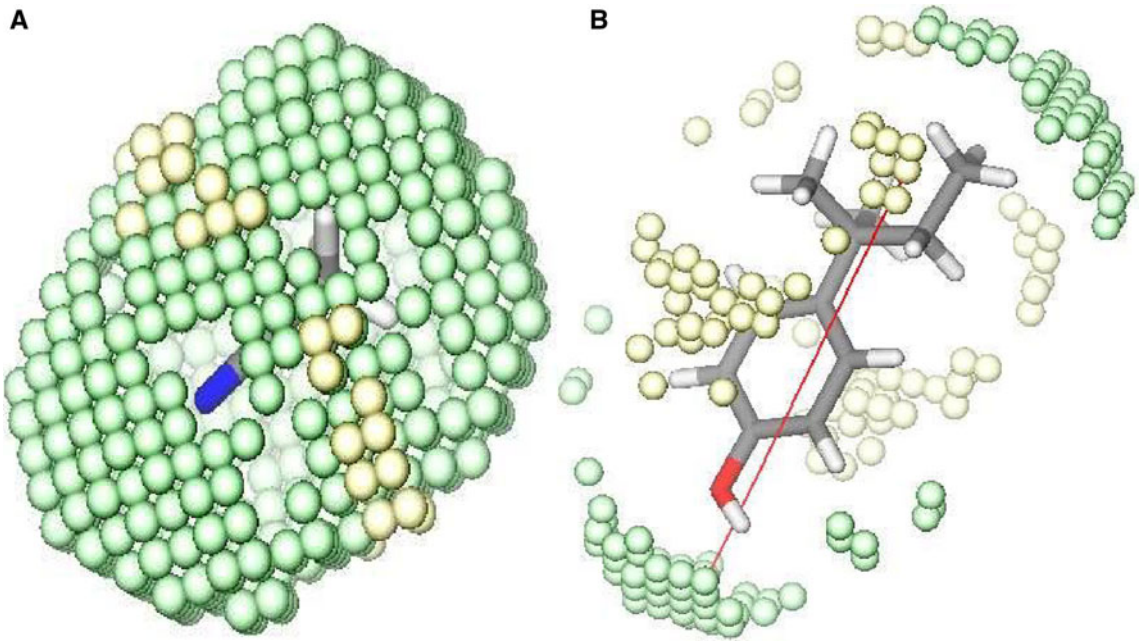

Fig. 4 Graphical display of GRIND variable 370 of the DRY-TIP cross-correlogram for (**a**): the most inactive (M5) and (**b**): the most active (M152) compound

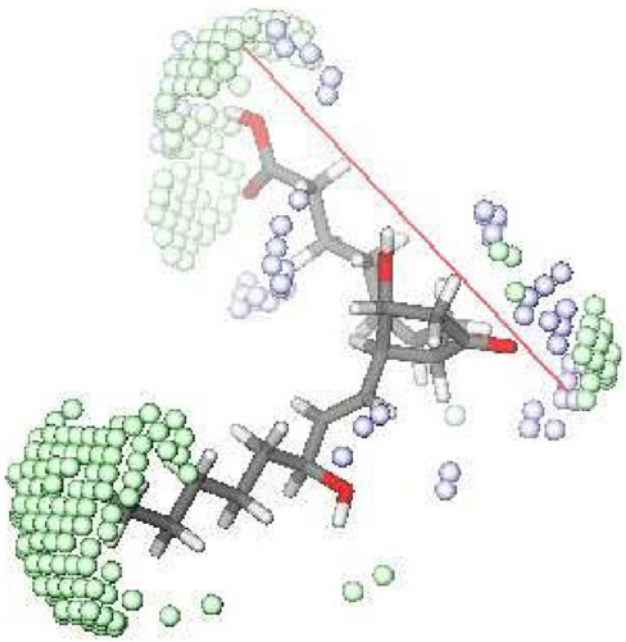

Fig. 5 Graphical display of GRIND variable 548 of the N1-TIP cross-correlogram of a selected (M233) compound

high affinity to aliphatic moieties. Therefore this variable is not expressed for the most inactive compound (Fig. 4). Variable 370 indicated a significant distance of 14–14.4 Å between TIP and N1 nodes which has positive correlation with stability constants vector. This variable is not expressed for the most inactive and most active compound.

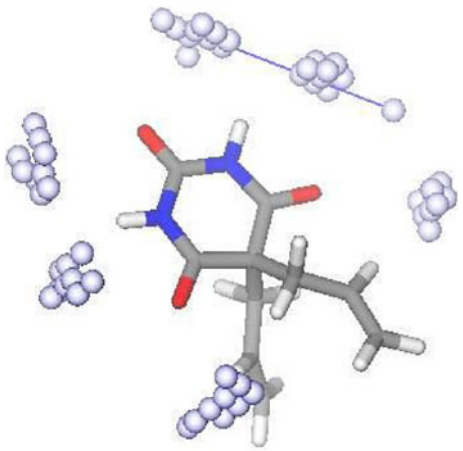

Fig. 6 Graphical display of GRIND variable 126 of the N1-N1 cross-correlogram of a selected (M135) compound

The graphical display of this variable for selected compound was shown in Fig. 5.

As N1 probe represents hydrogen bonding (HB) acceptor interaction, it is important in compounds with hydrogen bond donor group. The graphical display of this variable for selected compound was shown in Fig. 6. Variable 126, N1–N1 at distance of 4.8–5.2 Å, indicated an inverse relationship between this variable and stability constants. The negative impact on **y** observed in the auto-correlogram N1–N1 for this variable can be explained by an increasing polarity of guest molecules respect to non-polar cavity of CD. The most variables of N1–N1

correlogram that have negative correlation with **y** variable are located on the left side of the correlogram and variables with positively impact on the **y** variable are positioned on larger node–node distance. This indicated that spatial arrangement of HB regions of molecules is important in complexation of guest molecule to β-CD.

Conclusion

GRIND-based 3D QSAR models can give different kinds of information: simple and fast, reliable prediction of activity of compounds belonging to the data set and chemical interpretation of the results obtained. The main goal of this study was to investigate the reliability of Grind methodology in predicting stability constant of a huge and diverse class of guest molecules with β-CD. GRIND variables, TIP-TIP (6.8–7.2 Å) and TIP-DRY (11.2–11.6 Å), have high impact on stability constant. Strong relationship represented in TIP auto and cross correlogram indicated the importance of size-match of the host cavity to the guest, for inclusion complexation between guests molecules here investigated with β-CD. As earlier discussed, CDs complex formation usually results from different combinations of non-covalent interaction. Therefore all probes represent an impact on stability constant. Based on PLS coefficient, DRY probe, which represents hydrophobic interactions, had stronger impact on stability constant. It's concluded that steric and hydrophobic interactions are the mainly driving forces for β-CD complexation. The obtained 3D-QSAR which uses advanced technique of GRIND 3D-QSAR with better theoretical model is superior to previously reported models [13, 18] due to its combination of quality and speed.

References

1. Szejtli, J.: Introduction and general overview of cyclodextrin chemistry. Chem. Rev. **98**, 1743–1753 (1998)
2. Thompson, D.O.: Cyclodextrins-enabling excipients: their present and future use in pharmaceuticals. Criti. Rev. Ther. Drug. Carr. Syst. **14**, 1–104 (1997)
3. Hedges, A.H.: Industrial Applications of Cyclodextrins. Chem. Rev. **98**, 2035–2044 (1998)
4. Szejtli, J.: Past, present, and future of cyclodextrin research. Pure. App. Chem. **76**, 1825–1845 (2004)
5. Szejtli, J.: Utilization of cyclodextrins in industrial products andprocesses. J. Mater. Chem. **7**, 575–587 (1997)
6. Vyas, A., Saraf, S.: Cyclodextrin based novel drug delivery systems. J. Incl. Phenom. Macrocycl. Chem. **62**, 23–42 (2008)
7. Beck, M.T., Nagypal, I.: Chemistry of Complex Equilibria, Chap. 1. Horwood, Chichester. ISBN 0853121435. (1990)
8. Connors, K.A.: The Stability of Cyclodextrin Complexes in Solution. Chem. Rev. **97**, 1325–1357 (1997)
9. Loftsson, T., Stefansson, E.: Cyclodextrins in ocular drug delivery: theoretical basis with dexamethasone as a sample drug. J. Drug. Del. Sci. Tech. **17**, 3–9 (2007)
10. Lipkowitz, K.B.: Applications of Computational Chemistry to the Study of Cyclodextrins. Chem. Rev. **98**, 1829–1873 (1998)
11. Hansch, C., Leo, A.: Exploring QSAR: Fundamentals and Applications in Chemistry and Biology. American Chemical Society, Washington, DC (1995)
12. Estrada, E., Perdomo-Lopez, I., Torres-Labandeira, J.J.: Combination of 2D-, 3D-Connectivity and Quantum Chemical Descriptors in QSPR. Complexation of α- and β-Cyclodextrin with Benzene Derivatives. J. Chem. Inf. Comput. Sci. **41**, 1561–1568 (2001)
13. Perez-Garrido, A., Helguera, A.M., Guillen, A.A., Cordeiro, N.D.S., Escudero, A.G.: Convenient QSAR model for predicting the complexation of structurally diverse compounds with b-cyclodextrins. Bioorg. Med. Chem. **17**, 896–904 (2009)
14. Suzukia, T., Ishidaa, M., Fabian, W.M.F.: Classical QSAR and comparative molecular field analyses of the host-guest interaction of organic molecules with cyclodextrins. J. Comp. Aid. Mol. Des. **14**, 669–678 (2000)
15. Nayak, V.R., Kellogg, G.E.: Cyclodextrin-barbiturate inclusion complexes: A CoMFA/HINT 3-D QSAR study. Med. Chem. Res. **3**, 491–502 (1994)
16. Ghasemi, J.B., Saaidpour, S.: QSPR modeling of stability constants of diverse 15-crown-5 ethers complexes using best multiple linear regression. J. Incl. Phenom. Macrocycl. Chem. **60**, 339–351 (2008)
17. Ghasemi, J.B., Rofouei, M.K., Salahinejad, M.: A QSPR Study of the Stability Constant of Crown Ethers by Molecular Modelling: New Descriptors for Lariat Effect. J. Incl. Phenom. Macrocycl. Chem. In Press (2010)
18. Suzuki, T.: A Nonlinear Group Contribution Method for Predicting the Free Energies of Inclusion Complexation of Organic Molecules with α- and β-Cyclodextrins. J. Chem. Inf. Comp. Sci. **41**, 1266–1273 (2001)
19. Hudson, S.B., Hyde, M.R., Rahr, E., Wood, J.: Parameter based methods for compound selection from chemical databases. Quant. Struct. Act. Relat. **15**, 285–289 (1996)
20. Clark, R.D., Foz, P.C.: Statistical variation in progressive scrambling. J. Com. Aid. Mol. Des. **18**, 563–576 (2004)
21. Clark, R.D.: Boosted leave-many-out cross-validation: the effect of training and test set diversity on PLS statistics. J. Com. Aid. Mol. Des. **17**, 265–275 (2003)
22. Duran, A., Martinez, G.C., Pastor, M.: Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in Molecular Interaction Fields. J. Chem. Inf. Model. 48,1813-1823(2008); Pentacle 1.0.3; http://cadd.imim.es/gribcadd/projects/pentacle; http://www.moldiscovery.com (accessed on 29/09/2009)
23. Cratteri, P., Romanelli, M.N., Cruciani, G., Melani, C.B., Melani, F.: GRIND-derived pharmacophore model for a series of α-tropanyl derivative ligands of the sigma-2 receptor. J. Com. Aid. Mol. Des. **18**, 361–374 (2004)
24. Pastor, M., Cruciani, G., McLay, I., Pickett, S., Clementi, S.: GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. J. Med. Chem. **43**, 3233–3243 (2000)
25. Cruciani, G.: Molecular interaction fields, Applications in Drug Discovery and ADME Prediction, vol. 27. Wiley-VCH publication, Weinheim (2006)
26. Wold, S., Esbensen, K., Geladi, P.: Principal Component analysis. Chem. Intell. Lab. Syst. **2**, 37–52 (1987)
27. Dragos, H., Gilles, M., Alexandre, V.: Predicting the predictability: a unified approach to the applicability domain problem of QSAR models. J. Chem. Inf. Model. **49**, 1762–1776 (2009)

28. Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., Mekenyan, O.: A Stepwise Approach for Defining the Applicability Domain of SAR and QSAR Models. J.Chem. Inf. Model. **45**, 839–849 (2005)
29. Canceill, J., Lacombe, L., Collet, A.: Water-soluble cryptophane binding lipophilic guests in aqueous solution. J. Chem. Soc. Chem. Commun. **3**, 219–221 (1987)
30. Hirayama, F., Uekama, K.: In: Duchene, D. (ed.) Cyclodextrin and Their Industrial Uses, pp. 131–172. Editions De Sant′e, Paris (1987)