TÜBİTAK

Turkish Journal of Chemistry

http://journals.tubitak.gov.tr/chem/

Research Article

Turk J Chem
(2018) 42: 611 – 622

doi:10.3906/kim-1703-13

Ligand binding constants of the cucurbit[7]uril predicted with molecular docking: a theoretical study

Balázs ROÓSZ[1,*], **Tamás KÖRTVÉLYESI**[2,†], **Béla VISKOLCZ**[3]

[1]Doctoral School of Chemistry, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary
[2]Department of Physical Chemistry and Materials Science, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary
[3]Institute of Chemistry, Faculty of Materials Science and Engineering, University of Miskolc, Miskolc, Hungary

Received: 03.03.2017 • **Accepted/Published Online:** 14.08.2017 • **Final Version:** 01.06.2018

Abstract: A binding stability order predicting, molecular docking based fast technique was developed for host–guest complexes. Molecular descriptors were applied to ligand molecules to make the binding energy based docking scoring functions more efficient and reach the ±0.50 log K unit theoretical precision for predictions. The goal of this work was to model complexes of cucurbit[7]uril (CB7) as a host molecule with different local anesthetics and choline and phosphonium choline molecules. The guest molecules were docked into the cavity of CB7. The binding free energy maps were correlated by Contact, Grid-Based, Hawkins GB/SA, and AMBER docking scores obtained from UCSF DOCK software. The guest molecules were clustered by the scores. Linear correlations on the ln K - docking score plots were characterized on the learning set and log K was reproduced on the test set within an accuracy of ±0.92 log K units.

Key words: Thermodynamics, binding site, molecular recognition, host–guest, macrocycles, cucurbit[7]uril, molecular docking

1. Introduction

Host–guest complexes have a significant role in modern complex chemistry and drug development. The host molecules are macrocycles that can reversibly bind a smaller guest molecule. This phenomenon is widely used, for example in pharmaceutics to control the transport of the drug molecule to the corresponding receptor (controlled drug delivery: long acting local anesthetics, etc.).[1–3] In the following section the results are summarized, which underline the significance of host–guest chemistry in particular the in silico modeling of these systems. Kim et al. isolated and measured the X-ray crystal structure of three new CBn (n = 5, 7, and 8) homologues.[4] Hettiarachchi et al. measured "very low toxicity" for five CBn type molecules.[5] Uzunova et al. measured low toxicity for the CB7 and CB8 molecules in in vitro (CHO-K1 cell cultures, derived as a subclone from the parental Chinese hamster ovary -CHO- cell line) and in vivo (*Mus musculus*) experiments.[6] Khedkar et al. investigated CB7 complexes with the rhodamine B and kiton red S laser dyes with DFT, and found that the hydrogen bonds govern the stability of these systems.[7] Moghaddam et al. calculated ultrahigh affinity for the binding of the bicycle(2.2.2)octane and adamantane molecules in the CB7 cavity with the M2 affinity calculation method, and separated the binding entropies and enthalpies (Second Generation Mining Minima Software: using CHARMm molecular mechanics (MM) force field and generalized Born (GB) implicit solvation

*Correspondence: roosz@jgypk.szte.hu

model).[8] They developed the grid inhomogeneous solvation theory (GIST) to estimate solvation free energies and observed a toroidal region of high water density at the center of the CB7 host's cavity (these energetically and entropically disfavored water molecules contribute to the high host–guest stability constants of the CB7 molecule).[9] Gupta et al. analyzed the coumarin 1 molecule – CB7 complex with DFT and Hartree–Fock (HF) methods and showed that the interaction is "stabilized mainly by electrostatic, polarization and charge-transfer energy components".[10] Biedermann et al. investigated the high energy water molecules in the CB7 cavity by MD and DFT methods, and found that the release of these high energy water molecules causes very high guest affinity of the CBn macrocycles in aqueous solutions.[11] Sabet and Ganji modeled the CBn (n = 5–8) complexes of the fluorouracil drug using DFT methods with the Poisson–Boltzmann equation (PBE) and in addition performed ab initio MD simulations, finding that "... the dispersion interaction dramatically affects, as compared to the PBE based calculations, the binding of the host–guest complexes and is crucial to describe such systems".[12] Gavvala et al. investigated the interaction of the cardiotonic drug, milrinone, binding in the CB7 molecule with molecular docking (AutoDock software) and optimized the host–guest complex structure with a semiempirical (PM3MM) method.[13]

The field of CB7 host–guest chemistry became very exciting recently; many authors pointed out the essence of the "high energy" cavity waters and the importance of other above mentioned interactions during host–guest complex formation.[7,9–12] Studies showing low toxicity foresee the medical applicability of these systems (for drugs like anesthetics).[5,6] Wyman and Macartney obtained the host–guest stability constants (K_{CB7}) for five local anesthetics (procaine, tetracaine, procainamide, prilocaine, and dibucaine) and for 23 choline and phosphonium choline molecules using ^{1}H and ^{31}P NMR spectroscopies.[1,14] These stability constants were used for our molecular docking simulations (see Supplemental Information (SI), Table S1).

In this work the binding of guest molecules in cucurbit[7]uril (CB7) was studied by theoretical methods. The goal was the modeling of complex formation processes for the CB7 macrocycle, with molecular docking methods, and searching for a correlation between the calculated binding free energies (Gibbs free energies) and the experimental quantities (stability constants). A prediction method is suggested to find guest molecules with better binding properties.

2. Results and discussion

2.1. Molecular docking

First, some characteristic structural observations are summarized from the molecular docking and semiempirical quantum chemical geometry optimization results. In Figure 1 there is a typical docked structure from the [(procaine)H_2^{2+}] - CB7 (Novocain - CB7) inclusion complex. The docking results show that the acidic hydrogen atom of the protonated amine (ammonium) forms a H-bond with the oxo groups of the CB7 molecule. The aromatic rings bind to the hydrophobic cavity of the macrocycle. Using the AMBER scoring function, the guest molecule moved from the main spinning axle of the cucurbituril molecule and got closer to the wall of the cavity (see Figure 2). After the geometry optimization with the PM6 (Parameterization Method 6) method, the shape of the complex became elliptical (Figure 2), where the difference between the length of the two radii of ellipse varied from 0.2 to 2.0 Å (during the docking investigations, the host molecule was rigid and symmetric).

Applying the Grid-Based scoring function the guest molecules were divided into three groups (see Figure 3). These groups contain the following molecules (numbers from Table S1): (I., ●: A6, A7, A8, A9, A24, A25, A27, A28), (II., ■: A3, A4, A5, A12, A13, A14, A19, A20, A21, A22, A23, A26), (III., ▲: A1, A2, A10, A11,

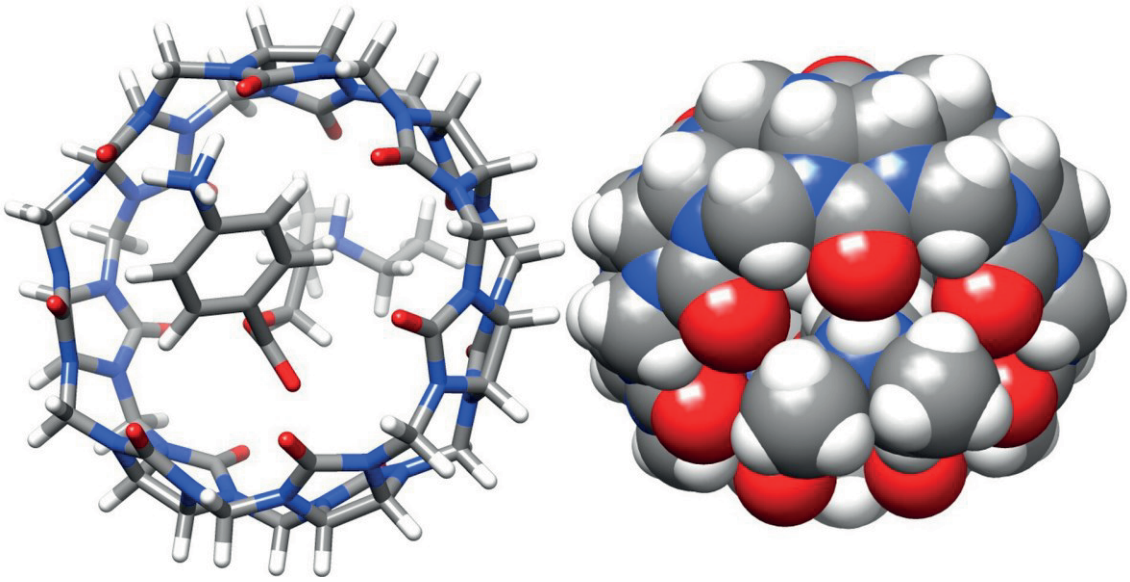

Figure 1. The host–guest complex of the $[(\text{procaine})\text{H}_2]^{2+}$ (Novocain) molecule in the CB7 molecule, docked with the Hawkins GB/SA scoring function. The protonated tertiary ammonium group and the protonated ammonium group form H-bonds in the two-opposite sides of the Novocain molecule.

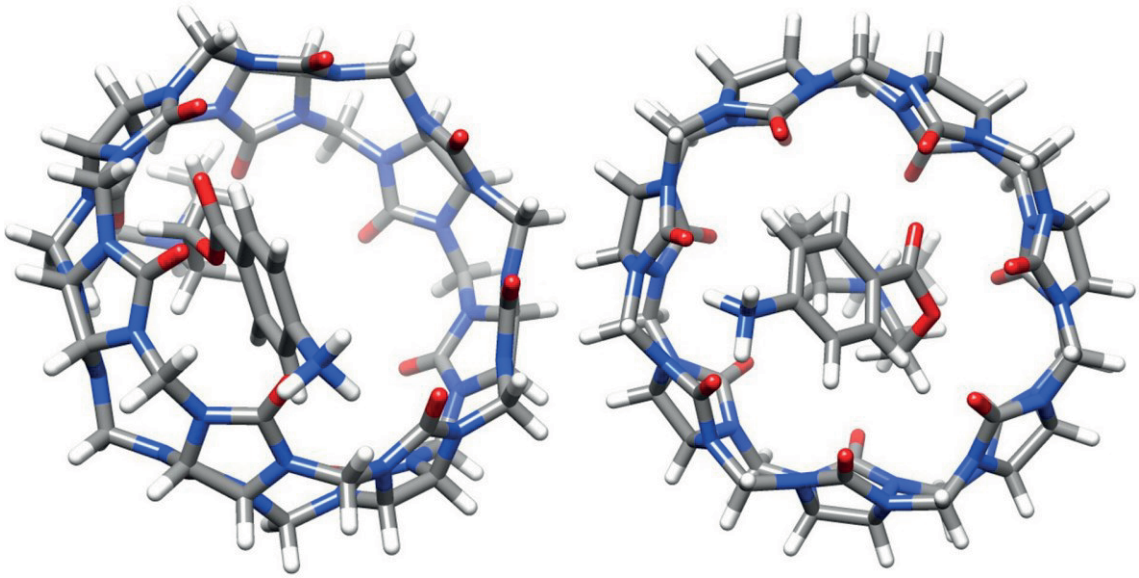

Figure 2. The complexes of the $[(\text{procaine})\text{H}_2]^{2+}$ molecule docked with the AMBER scoring function (left) and with the Hawkins GB/SA scoring function (right). After the docking the structures were optimized with the PM6 method and the CB7 structure showed a small deformation. During the AMBER docking process, the guest molecules moved from the center of the CB7 cavity towards the walls.

A15, A16, A17, A18). The Grid-Based score against the ln K shows linear correlation in all three cases (see SI: all statistical data from the fittings can be found in the SI data). The slopes of the linear equations are nearly equal, about $\sim 1.0 \times 10^{-4}$ (mol/J), and the entropy changes from Eq. (1) intercepts are $\Delta S^{\circ}_{I,}$ ■ = (–136.4 ±

49.0) J/(K mol), $\Delta S^{\circ}_{II,\,\bullet}$ = (–33.0 ± 28.5) J/(K mol), $\Delta S^{\circ}_{III,\,\blacktriangle}$ = (19.9 ± 21.8) J/(K mol) (Figure 3).

$$\left(\ln K = -\frac{\Delta H^{\circ}_{r}}{RT} + \frac{\Delta S^{\circ}_{r}}{R}\right) \tag{1}$$

The results of the learning set calculations are given in Table S2.

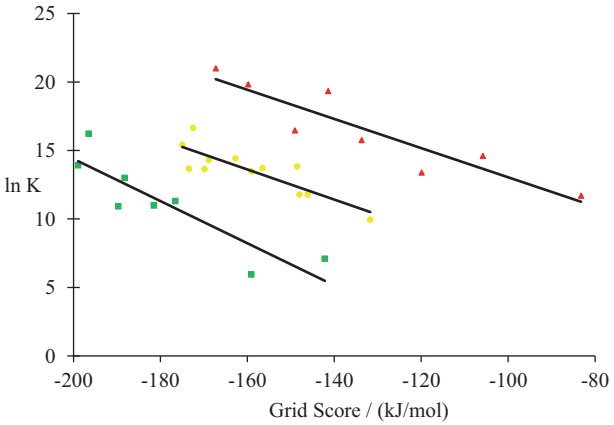

Figure 3. The logarithmic stability constant vs. the Grid-Based score, from the docking of molecules A1–A28 to the CB7 cavity. (I., ■: ln K = –(1.5392 ± 0.3276) × 10^{-4}× [Grid Score] – (16.4021 ± 5.8983); R^2 = 0.7863, II., ●: ln K = –(1.0987 ± 0.2146) × 10^{-4}× [Grid Score] – (3.9675 ± 3.4323); R^2 = 0.7238, III., ▲: ln K = –(1.0646 ± 0.1937) × 10^{-4} × [Grid Score] + (2.3953 ± 2.6188); R^2 = 0.8342).

Using the Hawkins GB/SA scoring function (generalized Born/surface area continuum solvation model), the results are similar to the values of the Grid-Based score. In this case the molecules were separated into three groups like with the Grid Score, but the guest number A26 is in group (I., ■) instead of group (II., ●), as Figure 4 shows. The slopes of the ln K vs. Hawkins GB/SA linear equations are about ∼1.0 × 10^{-4} (mol/J), the same value as in the case of the Grid-Based score function (Figure 3). The entropy changes from the intercepts are $\Delta S^{\circ}_{I,\,\blacksquare}$ = (–68.3 ± 24.0) J/(K mol), $\Delta S^{\circ}_{II,\,\bullet}$ = (11.1 ± 14.5) J/(K mol), $\Delta S^{\circ}_{III,\,\blacktriangle}$ = (50.2 ± 12.9) J/(K mol) (Figure 4). The linear equation fittings are slightly more accurate than the Grid-Based score (instead of 0.75 the R^2 is around 0.85).

The average slopes of the ln K vs. Hawkins GB/SA score and Grid-Based score linear equations are nearly equal within the deviation: (Grid-Based score slope = (–1.23 ± 0.25) × 10^{-4} (mol/J), a Hawkins GB/SA score slope = (–1.28 ± 0.18) × 10^{-4} (mol/J)). Moreover, it is in the same range with the 1/(8.314 J/(K mol) × 298.15 K) = –4.03 × 10^{-4} (mol/J) theoretical slope, which is from Eq. (1). Because of the same slopes, we can assume that the guest molecules bind the same way (hydrophobic guest molecule parts into the hydrophobic inner cavity of host molecule), and the difference between the intercepts is from the different entropy changes during the complex formation reactions. The unheeded main effect should be the displaceable water molecules from the cavity by a connecting guest causing huge solvation entropy, and the freezing torsion angles during complex formation in the hydrophobic cavity, which causes some conformation entropy changes too. However, the entropy effects of the binding in both cases could not be calculated explicitly during the docking simulations, and so three separated linear equations were fitted to the calculated energies. In the first

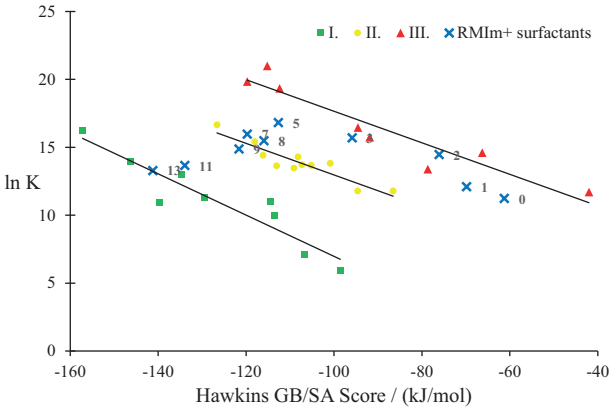

Figure 4. The logarithmic stability constant vs. the Hawkins GB/SA score, from the docking of molecules A1–A28 to the CB7 cavity. (I., ■: ln K = –(1.5187 ± 0.2251) × 10^{-4}× [Hawkins GB/SA Score] – (8.2182 ± 2.8838); R^2 = 0.8667, II., ●: ln K = –(1.1625 ± 0.1611) × 10^{-4}× [Hawkins GB/SA Score] + (1.3382 ± 1.7457); R^2 = 0.8525, III., ▲: ln K = –(1.1610 ± 0.1659) × 10^{-4}× [Hawkins GB/SA Score] + (6.0427 ± 1.5521); R^2 = 0.8908), ×: $RMIm^+$ surfactants, numbers show the flexible torsion angles.

approximation, the molecules in each regression group have the same extension (volume, surface, flexible torsion angles, etc.) and so the different regression groups are dissimilar in their molecular extension parameters. The 1-alkyl-3-methylimidazolium ($RMIm^+$) surfactant structures from the work by Miskolczy et al. were used to show the effect of flexible torsion angles and molecular extension.[15] The surfactants with 11 and 13 flexible torsion angles are in group I. ■ (learning set molecules with 4–11 flex. torsion angles), 7–9 torsion angles in group II. ● (4–7 flex. torsion angles), and 0–3 torsion angles in group III. ▲ (0–4 flex. torsion angles) (see Figure 4). The docking results of surfactant molecules show that the number of flexible torsion angles is an important molecular descriptor (usually growing with the molecular extension), but not solely responsible for the interactions. To reach the correct model and separate the guest molecules by entropy changes, additional molecular descriptors were used. These are the following quantities: $A_{(DMS)}$: vdW surface area of the guest molecule calculated with the DMS software, $V_{(COSMO)}$: COSMO volume (from SAS: solvent accessible surface area an excluded volume calculated for the COSMO model), $V_{(in\ cavity)}$: volume of the guest molecule in the cavity, $n_{(flexible\ torsion)}$: the number of flexible torsions in the guest molecule. Four molecular descriptors show an inverse relationship with the intercepts (entropy changes), as represented in Figure 5.

The number of displaced "high energy" cavity water molecules increases with $V_{(in\ cavity)}$, and so the solvation entropy change is increased with this quantity (read the end of the SI data: many other molecular descriptors are identified by a short QSAR study that increases the number of displaced cavity waters). When the number of flexible torsion angles increases, the conformation entropy change decreases, because the degrees of freedom increase with the number of flexible torsions and a very flexible molecule can lose plenty of degrees of freedom during host–guest complex formation. This in turn increases the Gibbs free energy and lowers log K. The small, rigid molecules have positive entropy change (entropy controlled binding, relatively small enthalpy, few H-bonds, and weak vdW interactions) due to the few frozen torsion angles and the relatively high solvation entropy (hydrate layer loss and displaced cavity water molecules). On the other hand, the large, flexible molecules with many rotatable bonds have enthalpy controlled binding (more H-bonds and stronger

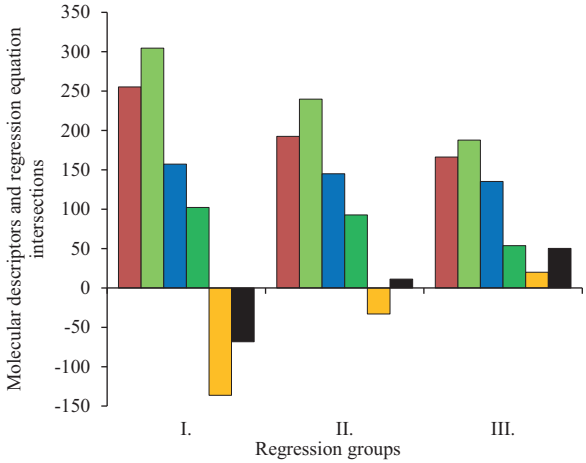

Figure 5. The molecular descriptors and the entropy changes (Hawkins GB/SA score). The examined quantities are in reverse rate with the $\Delta S^{\circ}_{(Grid-Based\,score)}$ and the $\Delta S^{\circ}_{(Hawkins\,GB/SA\,score)}$ entropies. This can explain the three regression groups (I., II., and III.). ■: $A_{(DMS)}$ / Å^2, ■: $V_{(COSMO)}$ / Å^3, ■: $V_{(in\,cavity)}$ / Å^3, ■: 10 × $n_{(flexible\,torsions)}$ / 10 × pieces of torsions, ■: $\Delta S^{\circ}_{(Grid\,Based\,score)}$ / J/ (K mol), ■: $\Delta S^{\circ}_{(Hawkins\,GB/SA\,score)}$ / J/ (K mol). The colors are in the same order in the figure and the caption.

vdW interactions than small molecules). The solvation entropy increases with the number of displaced "high energy" host cavity water molecules (and lost ligand solvent shell). The conformational entropy change decreases (more negative value) with the increasing number of flexible ligand torsion angles. These are inverse effects, which compete with each other. The most stable CB7 host–guest complexes have totally rigid guest molecules, which displace the maximum number of "high energy" waters from the host cavity, and have a suitable apolar segment, which can bind to the apolar cavity and maybe hold flexible H donor side chains for H-bonds with oxonium groups.

In the case of the Grid-Based score, the relative entropy loss between group (II., ●) and (I., ■) was $\Delta S^{\circ}_{rel.}$ = –103.4 J/(K mol) and between the (III., ▲) and (II., ●) groups it was $\Delta S^{\circ}_{rel.}$ = –52.9 J/(K mol). Using the Hawkins GB/SA score the corresponding quantities were the following: between group (II., ●) and (I., ■) $\Delta S^{\circ}_{rel.}$ = –79.5 J/(K mol) and between (III., ▲) and (II., ●) $\Delta S^{\circ}_{rel.}$ = –39.1 J/(K mol). In both cases the quotient of the two entropy losses is nearly an integer, namely two. Comparing the difference among groups I–II and II–III, Eq. (2) may show twice as much cavity waters leaving between groups I and II than II and III. This approximating statement should be valid if the enthalpy changes of the equilibria reactions (with water displacement number difference not greater than 1 or 2) are nearly the same.

$$(\Delta S^{\circ}_{rel.\ (I.,\blacksquare - II.,\bullet)} / \Delta S^{\circ}_{rel.\ (II.,\bullet - III.,\blacktriangle)}) \approx 2.0 \tag{2}$$

Applying the AMBER score the molecules were divided into four groups (A, □; B, ◇; C, ○; D, △) by the ln K vs. AMBER score linear regressions (see Figure 6). The groups contain the following guest molecules: (A, □: A7, A25, A26, A27), (B, ◇: A3, A6, A8, A28), (C, ○: A1, A2, A4, A5, A12, A13, A14, A24), (D, △: A10, A11, A15, A18, A19, A20, A21, A22, A23). The molecules no. A9 (absent force field parameters), A16, and A17 (do not fit into any groups) are missing. The AMBER and the Hawkins GB/SA regression groups

do not match each other completely (plots in Figure 6 are colored by the Hawkins GB/SA regression groups); the AMBER score makes the data set more diverse than Hawkins GB/SA score because of the more complex scoring function. Applying the GB solvation model (Figure 6) showed that the slopes of the linear equations are nearly equal (slope = -1×10^{-1}), showing a slow decrease in the D, C, B, A order. The intercepts decrease in the same order as the previous cases. The molecules in each group have similar surface areas, volumes, and filled volumes in the CB7 cavity. The extension of the molecules grows in the D, C, B, A order, but the average surface area and volume changes are small between the A–B and B–C group pairs. Figure 7 shows the molecular descriptors and the entropy changes (intercepts of AMBER score fitted linear equations).

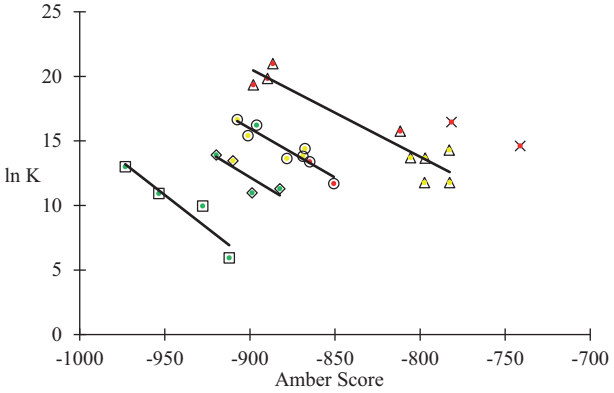

Figure 6. The logarithmic stability constant vs. the AMBER Score, Hawkins, Cramer, and Truhlar GB solvation model, from the docking of molecules A1–A28 to the CB7 cavity. The green, yellow, and red colors show the regression groups (green: I., yellow: II., red: III.) obtained from the Hawkins GB/SA vs. ln K fitting. The AMBER score makes the results more diverse and the regression groups overlap each other. (A: □; ln K = $-(1.0260 \pm 0.2660) \times 10^{-1} \times$ [Amber Score] – (86.6670 ± 25.0644); $R^2 = 0.8814$, B: ◇; ln K = $-(7.9810 \pm 3.3674) \times 10^{-2} \times$ [Amber Score] – (59.6434 ± 30.4033); $R^2 = 0.7374$, C: ○; ln K = $-(7.6047 \pm 1.1759) \times 10^{-2} \times$ [Amber Score] – (52.4759 ± 10.3430); $R^2 = 0.8745$, D: ; ln K = $-(6.8172 \pm 0.9277) \times 10^{-2} \times$ [Amber Score] – (40.7651 ± 7.6928); $R^2 = 0.8852$).

2.2. Log K prediction

The investigated complexes and stability constants are from different origins, and so the deviations of the stability constants are different as well. The docking results of these molecules are collected in Table S3. To predict the stability constants, the previously calculated ln K – Hawkins GB/SA score linear equations were used. To find the best regression group (I., ■, II., ●, III., ▲) for each molecule (where the ln K – Hawkins GB/SA equation is the most valid for the corresponding molecule), linear regression was applied with four molecular descriptors on the "learning set" (A1–A28) molecules. The variables were: the Grid-Based score, the DMS surface area of the guest molecule, the number of the rotatable torsion angles, and the Contact Score (the number of the ligand–receptor atom pairs closer than 4.5 Å). The dependent variable was the Hawkins GB/SA score. With the well-known choline and anesthetic guest molecules (A1–A28), linear equations were fitted to identify the regression groups (Hawkins GB/SA score vs. the four descriptors, Table) for test set molecules.

The Hawkins GB/SA score was calculated with these equations (Table) for the guest molecules in Table S4. Using these "predicted" Hawkins GB/SA scores and the calculated GB/SA scores (from docking), the (I., ■, II., ●, III., ▲) regression groups were identified from the absolute error of the two scores for the test

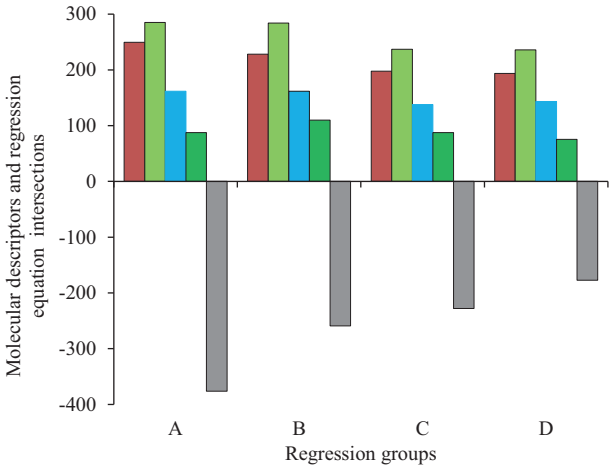

Figure 7. The molecular descriptors and the entropy changes (AMBER score). The AMBER score makes the system too diverse, because we have not got enough data points. ■: $A_{(DMS)}$ / Å^2, ■: $V_{(COSMO)}$ / Å^3, ■: $V_{(in\,cavity)}$ / Å^3, ■: 10 × $n_{(flexible\,torsions)}$ / 10 × pieces of torsions, ■ 10 × regression $intercept_{(AMBER\,score)}$. The colors are in the same order in the figure and the caption.

Table. Linear regressions for the identification of regression groups. To the "learning set", molecules A1–A28, we fitted these equations and used these equations to find the regression groups on test set molecules 29–52, to predict the stability constants.

Group	Slope$_{"Grid-Based"}$	Slope$_{surf.\,A.}$	Slope$_{flex.\,tors.}$	Slope$_{"Contact\,Score"}$	Intercept	R^2
I.	0.4920	–0.04311	0.7816	0.04469	–1.3047	0.87
II.	1.2096	0.0009798	1.0847	–0.003687	10.3230	0.77
III.	0.3811	–0.07380	0.2870	0.07893	8.5936	0.97

molecules. Table S4 shows the identification of regression groups. The smallest absolute error from the docked GB/SA score belongs to the gray cells. These cells were used to identify the regression groups (I., II., III.) for guest molecules. Knowing the predicted regression groups the binding constants were estimated with the equations (Figure 4). In Table S5 from the SI, the gray cells display the best calculated stability constants (these log K's are closest to the experimental one). The "calculated best log K" column represents the best log K's (closest to the experimental log K) selected from calculated ones by the three regression groups (in this case the average absolute error was ±0.48 log K units). The "predicted log K" column shows those stability constants that were calculated using the identified regression groups (the average absolute error was ±0.92 log K units). The maximum theoretical precision of this model (the entropy loss was not explicitly modelled) could be about ±0.5 log K units, and the actual precision is near ±0.9 log K units. Investigating the test set only 9 molecules have more than 3 flexible torsion angles (4–8, mainly 6) and 15 molecules have 3 or less than 3 flexible torsion angles; these are rigid guest molecules. The rigid molecules during the docking study may misguide the prediction results, because in the original anchor-and-grow algorithm validation article 79% of crystal ligand poses were reproduced within 2 Å error using the rigid ligand docking algorithm and 72% were reproduced applying the flexible ligand docking algorithm.[15] This effect was investigated and it was found that the average absolute errors of predicted log K's are nearly the same using the flexible (±0.94) or rigid (±0.90)

test set molecules during our prediction. The explanation should be that in the CB7-guest molecule complexes the solvation entropy effects (mainly from displaced cavity waters) are more dominant than the conformation entropy part. Attempts were made to model the entropy changes (number of displaced water molecules), with QSAR study chosen few molecular descriptors (see SI).

The binding stability order was predicted for guest molecules by using manually selected simple molecular descriptors (surface area, number of flexible torsion angles, etc.) and the Hawkins GB/SA scoring function. This can be a very beneficial solution for forecasting the equilibrium data (log K order) of these systems, which requires a lot of work to map (for example with competitive experiments or expensive computations or see the efforts of Muddana et al. based on their self-developed M2 force-field based method).[17,18] The predicted equilibrium constants are useful for pharmaceutical and analytical applications and to design ultrahigh affinity host–guest complexes.[19]

2.3. Conclusions

During this work, anesthetic molecules and choline derivatives were docked to the CB7 host molecule. As indicated, the guest molecules bind in the same way to the macrocycle. The apolar function groups bind to the hydrophobic cavity while the polar groups, like the primary, secondary, tertiary amine, quaternary ammonium, or phosphonium groups, bind near the polar oxo-groups. This is confirmed by the similar slopes of the stability constant – Hawkins GB/SA score regressions. The AMBER force-field based dockings showed that the guest molecule binds near the wall of the macrocycle cavity, and not in the vertical symmetry axle. From the semiempirical quantum chemistry (PM6) calculations it is evident, that the macrocycle deforms during the binding and does not remain symmetric. In the ln K vs. docking score regressions more than one linear equation was obtained with different intercepts, and this can be explained by the fact that the entropy change was not calculated in our models. In the case of molecules with few or many rotatable torsion angles the differences between the intercepts are mainly from the displaced "high energy" water molecules and the solvation entropy changes of ligands. The minor effect is from the loss of the degrees of freedom during complex formation (frozen torsion angles cause entropy loss). Using the Hawkins GB/SA scoring function, this relative entropy change is $\Delta S^{\circ}_{rel.}$ = –79.5 J/(K mol) between the largest and the middle-sized molecules and $\Delta S^{\circ}_{rel.}$ = –39.1 J/(K mol) between the middle-sized and the smallest guest molecules. Both with the Grid-Based and the Hawkins GB/SA scoring functions, the ratio of the two entropy losses is 2/1, respectively (–79.5/–39.1 = 2.03) (Eq. (2)). This ratio may indicate that twice as many water molecules leave between the largest and middle-sized than between the middle-sized and smallest guest molecules. An inverse proportion was found for the average surface area, average molecule volume, average number of rotatable torsion angles, and between the entropy changes of the reaction in the regression groups. When the above written descriptors increase, the entropy change decreases (the entropy loss increases). It could be possible to predict the entropy change for molecules with unknown experimental stability constants by a QSAR study predicting the number of leaving cavity water molecules from appropriate molecular descriptors, and estimating the log K from the previously fitted experimental log K – docking score (Hawkins GB/SA) equations. This can be a very fast method to show the binding stability order of different guest molecules with a host molecule. Using the molecular descriptors of the ligand molecules makes the binding energy based docking scoring functions more effective. The maximum theoretical precision of our model could be about ±0.48 log K units, and the actual precision is ±0.92 log K units.

3. Experimental

The structure of the CB7 molecule was optimized with Gaussian 03 in the BHAND&HLYP/6-31G* level.[20] The structures of the guest molecules were optimized with the PM6 semiempirical quantum chemical method implemented in the MOPAC2012 package.[21] Docking of guest molecules on the host molecule was carried out by the UCSF DOCK 6.3 software applying the anchor-and-grow algorithm.[16,22,23] The molecules were prepared for docking with these program packages: the molecular surface defined by Richards was calculated with the DMS method and net atomic charges were calculated with the AM1-BCC method, implemented in the Antechamber software for host and guest molecules.[24,25] During the docking procedure the primary score was the Grid-Based Score and the secondary score was the Hawkins GB/SA Score.[26] The docked structures were optimized with the PM6 method and the molecular volumes ($V_{(COSMO)}$: COSMO volume) were given from the COSMO solvation model (COnductor-like Screening MOdel).[27] The docked conformations (from the Hawkins GB/SA Score) were redocked with the AMBER molecular mechanics MM-GBSA score based on the AMBER molecular mechanics force field.[28] During the force-field docking the GAFF (general AMBER) force-field was used with the Hawkins, Cramer, and Truhlar pairwise GB model, parametrized by Tsui and Case.[29,30] This is a generalized Born (GB) implicit solvation method. The calculated Gibbs free energy score values were used to give the binding energies.

During the geometry optimization with the PM6 method, we applied the EF (Eigenvector Following) minimization routine. Under the docking approach with the Grid Based and Hawkins GB/SA docking functions the maximum number of starting molecule segments was 20,000, the minimal number of heavy atoms in the starting molecule segments was 4, and the maximum number of anchor orientations carried forward in the anchor-and-grow search was 5000; the pruning value cutoff for anchor orientations promoted to the conformational search was 100. With the Grid-Based Score, the distance dependent dielectric implicit water model was used (charge – charge distance (r) dependent dielectric, $\varepsilon = 4 \times r$). The dielectric constant (ε) of the water for the GB/SA Hawkins solvation model was 78.5 and the salt screening for Hawkins GB/SA score was applied with 0.1 mol/dm^3 salt concentration and 0.09 GB radius offset. In the case of the AMBER score we employed 1000 steps of pre- and post-minimization and 10,000 steps of Langevin molecular dynamics (MD). After Grid-Based and Hawkins GB/SA based docking the bumped docked conformations were removed by bump filter (all the best docked conformations were checked by hand and the remaining clashing ones were thrown away). The Grid-Based scoring function contains electrostatic and vdW interactions, the solvation energy calculated with dielectric term that increases with distance ($4 \times r$). The Hawkins GB/SA scoring function contains a GB/SA solvation term (in addition to the Coulomb and vdW terms of the Grid-Based Score) that was parametrized to approach the free energy of hydration. During the molecular docking and prediction study, the molecules no. A1–A28 were the "learning set" and molecules no. 29–52 were the "test set". The stability constants (K_{CB7}) of the learning and test set are from the reference articles in Table S1.[1,14,31–42] The 1-alkyl-3-methylimidazolium ($RMIm^+$) surfactant structures were prepared for molecular docking with the above described (PM6 geom. opt., AM1-BCC charges, Hawkins GB/SA Score docking) protocol.[15] The UCSF Chimera software was used for the molecular structure visualization.[43]

Acknowledgments

We are grateful to the Hungarian Supercomputer Centre, NIIF and HPC Szeged for the computational possibilities and the projects named "TÁMOP-4.2.2.C-11/1/KONV-2012-0010 – Supercomputer, national virtual

laboratory" and "TÁMOP-4.1.1.C-12/1/KONV-2012-0014" supported by the European Union and co-financed by the European Social Fund. Special thanks to Lily Hunnisett, MSc Chemist (Queen Mary University of London, United Kingdom) for the grammar corrections!

References

1. Wyman, I. W.; Macartney, D. H. *Org. Biomol. Chem.* **2010**, *8*, 247-252.
2. Grant, S. A. *Best Pr. Res. Clin. Anaesthesiol.* **2002**, *16*, 345-352.
3. Lagona, J.; Mukhopadhyay, P.; Chakrabarti, S.; Isaacs, L. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 4844-4870.
4. Kim, J.; Jung, I. S.; Kim, S. Y.; Lee, E.; Kang, J. K.; Sakamoto, S.; Yamaguchi, K.; Kim, K. *J. Am. Chem. Soc.* **2000**, *122*, 540-541.
5. Hettiarachchi, G.; Nguyen, D.; Wu, J.; Lucas, D.; Ma, D.; Isaacs, L.; Briken, V. *PLoS One* **2010**, *5*, e10514.
6. Uzunova, V. D.; Cullinane, C.; Brix, K.; Nau, W. M.; Day, A. I. *Org. Biomol. Chem.* **2010**, *8*, 2037-2042.
7. Khedkar, J. K.; Jagtap, K. K.; Pinjari, R. V; Ray, A. K.; Gejji, S. P. *J. Mol. Model.* **2012**, *18*, 3743-3750.
8. Moghaddam, S.; Yang, C.; Rekharsky, M.; Ko, Y. H.; Kim, K.; Inoue, Y.; Gilson, M. K. *J. Am. Chem. Soc.* **2011**, *133*, 3570-3581.
9. Nguyen, C. N.; Young, T. K.; Gilson, M. K. *J. Chem. Phys.* **2012**, *137*, 44101.
10. Gupta, M.; Maity, D. K.; Singh, M. K.; Nayak, S. K.; Ray, A. K. *J. Phys. Chem. B* **2012**, *116*, 5551-5558.
11. Biedermann, F.; Uzunova, V. D.; Scherman, O. A.; Nau, W. M.; De Simone, A. *J. Am. Chem. Soc.* **2012**, *134*, 15318-15323.
12. Sabet, M.; Ganji, M. D. *J. Mol. Model.* **2013**, *19*, 4013-4023.
13. Gavvala, K.; Koninti, R. K.; Sengupta, A.; Hazra, P. *Phys. Chem. Chem. Phys.* **2014**, *16*, 2823-2826.
14. Wyman, I. W.; Macartney, D. H. *Org. Biomol. Chem.* **2010**, *8*, 253-260.
15. Miskolczy, Z.; Biczók, L.; Megyesi, M.; Jablonkai, I. *J. Phys. Chem. B* **2009**, *113*, 1645-1651.
16. Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. *J. Comput. Aided. Mol. Des.* **2006**, *20*, 601-619.
17. Muddana, H.; Fenley, A.; Mobley, D.; Gilson, M. *J. Comput. Aided. Mol. Des.* **2014**, *28*, 305-317.
18. Muddana, H.; Yin, J.; Sapra, N.; Fenley, A.; Gilson, M. *J. Comput. Aided. Mol. Des.* **2014**, *28*, 463-474.
19. Shetty, D.; Khedkar, J. K.; Park, K. M.; Kim, K. *Chem. Soc. Rev.* **2015**, *44*, 8747-8761.
20. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Fox, D. J.; et al. Gaussian 03, Revision D.02 Gaussian, Inc.: Wallingford, CT, USA, 2004.
21. Stewart, J. J. *J. Mol. Model.* **2007**, *13*, 1173-1213.
22. Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. *RNA* **2009**, *15*, 1219-1230.
23. Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput. Aided. Mol. Des.* **2001**, *15*, 411-428.
24. Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151-176.
25. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132-146.
26. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122-129.
27. Klamt, A.; Schuurmann, G. *J. Chem. Soc. Perkin Trans. 2* **1993**, No. 5, 799-805.

28. Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. *J. Mol. Biol.* **2008**, *377*, 914-934.

29. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

30. Tsui, V.; Case, D. A. *Biopolymers* **2000**, *56*, 275-291.

31. Zhou, Y. Y.; Yang, J.; Liu, M.; Wang, S. F.; Lu, Q. *J. Lumin.* **2010**, *130*, 817-820.

32. Miskolczy, Z.; Biczók, L.; Görner, H. *J. Photochem. Photobiol. A Chem.* **2009**, *207*, 47-51.

33. Wang, R.; Macartney, D. H. *Tetrahedron Lett.* **2008**, *49*, 311-314.

34. Petrov, N. K.; Ivanov, D. A.; Golubkov, D. V; Gromov, S. P.; Alfimov, M. V. *Chem. Phys. Lett.* **2009**, *480*, 96-99.

35. Yuan, L.; Wang, R.; Macartney, D. H. *Tetrahedron: Asymmetry* **2007**, *18*, 483-487.

36. Li, C. F.; Du, L. M.; Zhang, H. M. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2010**, *75*, 912-917.

37. Li, C. F.; Du, L. M.; Wu, W. Y.; Sheng, A. Z. *Talanta* **2010**, *80*, 1939-1944.

38. Li, Y. P.; Wu, H.; Du, L. M. *Chinese Chem. Lett.* **2009**, *20*, 322-325.

39. Wang, R.; Wyman, I.; Wang, S.; Macartney, D. *J. Incl. Phenom. Macrocycl. Chem.* **2009**, *64*, 233-237.

40. Wang, R.; Yuan, L.; Macartney, D. H. *Chem. Commun.* **2005**, No. 47, 5867-5869.

41. Montes-Navajas, P.; Garcia, H. *J. Photochem. Photobiol. A Chem.* **2009**, *204*, 97-101.

42. Liu, S.; Ruspic, C.; Mukhopadhyay, P.; Chakrabarti, S.; Zavalij, P. Y.; Isaacs, L. *J. Am. Chem. Soc.* **2005**, *127*, 15959-15967.

43. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605-1612.

Supplemental Information (SI)

Table S1. Guest molecules studied in the calculations, their stability constants. and the number of the flexible torsion angles. Molecules A1–A28 are the learning set and 29–52 are the test set for the stability constant prediction. The green, yellow, red color code shows the ln K – Hawkins GB/SA score regression group (articles are from main Ref.).

No.	"Guest" molecule	$K_{CB[7]}$, $dm^3\ mol^{-1}$	$N_{flexible\ torsions}$, pcs.	Group	Article
A1	$N^+(CH_3)_4$	$(1.2 \pm 0.4) \times 10^5$	4	III.	14
A2	$(CH_3)_3N^+(CH_2)_2OH$	$(6.5 \pm 1.2) \times 10^5$	6	III.	14
A3	$(CH_3)_3N^+(CH_2)_2O_2CCH_3$	$(7.0 \pm 1.3) \times 10^5$	8	II.	14
A4	$(CH_3)_3N^+CH_2CH(CH_3)O_2CCH_3$	$(4.9 \pm 0.9) \times 10^6$	9	II.	14
A5	$(CH_3)_3N^+(CH_2)_2O_2C(CH_2)_2CH_3$	$(1.7 \pm 0.3) \times 10^7$	10	II.	14
A6	$(CH_3)_3N^+(CH_2)_{11}CH_3$	$(5.8 \pm 0.1) \times 10^4$	15	I.	14
A7	$(CH_3)_3N^+CH_2CH(OH)CH_2CO_2^-$	$(3.8 \pm 0.5) \times 10^2$	8	I.	14
A8	$(CH_3)_3N^+CH_2CH(OH)CH_2CO_2H$	$(8.0 \pm 1.1) \times 10^4$	9	I.	14
A9	$(CH_3)_3N^+(CH_2)_2OP(O)(OH)O^-$	$(1.2 \pm 0.2) \times 10^3$	8	I.	14
A10	$(PhCH_2)N^+(CH_3)_3$	$(2.5 \pm 0.2) \times 10^8$	5	III.	14
A11	$(PhCH_2)(CH_3)_2N^+(CH_2)_2OH$	$(4.1 \pm 1.2) \times 10^8$	7	III.	14
A12	$N^+(CH_2CH_3)_4$	$(1.0 \pm 0.2) \times 10^6$	8	II.	14
A13	$(CH_3CH_2)_3N^+(CH_2)_2OH$	$(1.8 \pm 0.3) \times 10^6$	9	II.	14
A14	$(CH_3CH_2)_3N^+(CH_2)_4CH_3$	$(8.3 \pm 1.9) \times 10^5$	11	II.	14
A15	$(quinuclidine)^+(CH_2)_2OH$	$(1.3 \pm 0.3) \times 10^9$	3	III.	14
A16	$P^+(CH_3)_4$	$(2.2 \pm 0.4) \times 10^6$	4	III.	14
A17	$(CH_3)_3P^+(CH_2)_4CH_3$	$(1.4 \pm 0.3) \times 10^7$	8	III.	14
A18	$(CH_3)_3P^+(CH_2)_2OH$	$(6.9 \pm 1.8) \times 10^6$	6	III.	14
A19	$(CH_3)_3P^+(CH_2)_2O_2CCH_3$	$(8.9 \pm 1.7) \times 10^5$	8	II.	14
A20	$P^+(CH_2CH_3)_4$	$(1.3 \pm 0.3) \times 10^5$	8	II.	14
A21	$(CH_3CH_2)_3P^+(CH_2)_4CH_3$	$(1.6 \pm 0.3) \times 10^6$	11	II.	14
A22	$(CH_3CH_2)_3P^+(CH_2)_2OH$	$(1.3 \pm 0.3) \times 10^5$	9	II.	14
A23	$(CH_3CH_2)_3P^+(CH_2)_2O_2CCH_3$	$(8.6 \pm 1.6) \times 10^5$	11	II.	14
A24	$(dibucaine)H_2^{2+}$	$(1.1 \pm 0.2) \times 10^7$	13	I.	1
A25	$(procainamide)H_2^{2+}$	$(5.5 \pm 1.1) \times 10^4$	9	I.	1
A26	$(prilocaine)H^+$	$(2.1 \pm 0.4) \times 10^4$	8	I.	1
A27	$(procaine)H_2^{2+}$	$(4.4 \pm 1.6) \times 10^5$	10	I.	1
A28	$(tetracaine)H_2^{2+}$	$(1.1 \pm 0.3) \times 10^6$	12	I.	1
29	melamine	$(4.70 \pm 0.12) \times 10^4$	3		30
30	lumichrome, (Lc)	$(8.32 \pm 1.15) \times 10^3$	2		31
31	BDC^+	$(2.0 \pm 0.5) \times 10^4$	8		32
32	$DTCI^+$	$(2.8 \pm 0.2) \times 10^4$	4		33
33	$BNEA(H)^+$	$(1.05 \pm 0.13) \times 10^8$	5		34
34	$sanguinarine^+$, (SA^+)	4.53×10^4	1		35
35	$coptisine^+$	1.86×10^4	0		36
36	$berberine^+$, (BR^+)	9.57×10^4	4		37
37	$norharmane(H)^+$, $(NHMH^+)$	$(9.0 \pm 0.5) \times 10^4$	0		38
38	$2\text{-}aminoanthracene(H)^+$	8.0×10^5	1		39
39	$profalvine(H_2)^{2+}$	$(1.70 \pm 0.25) \times 10^7$	2		40
40	$pyronin(H)^{2+}$	$(1.47 \pm 0.48) \times 10^7$	2		40
41	$oxonine(H)^{2+}$	$(1.21 \pm 0.16) \times 10^7$	2		40
42	$tionine(H)^{2+}$	$(1.35 \pm 0.10) \times 10^7$	2		40
43	$acridine\text{-}orange(H_2)^{2+}$	$(0.31 \pm 0.05) \times 10^7$	6		40
44	$pyronine\text{-}Y(H)^+$	$(0.46 \pm 0.01) \times 10^7$	6		40
45	$methylene\ blue(H)^+$	$(1.26 \pm 0.28) \times 10^7$	6		40
46	$profalvine(H)^+$	$(2.35 \pm 0.57) \times 10^6$	2		40
47	$pyronine^+$	$(3.40 \pm 1.41) \times 10^6$	2		40
48	$oxonine^+$	$(5.44 \pm 0.72) \times 10^6$	2		40
49	$tionine^+$	$(3.05 \pm 0.53) \times 10^6$	2		40
50	$acridine\text{-}orange(H)^+$	$(0.42 \pm 0.07) \times 10^6$	6		40
51	$pyronine\text{-}Y^+$	$(2.01 \pm 0.34) \times 10^6$	6		40
52	$1,2\text{-}phenylenediamine(H_2)^{2+}$	$(8.04 \pm 1.28) \times 10^4$	2		41

Table S2. Properties of the anesthetic and choline molecules. The Grid-Based score (kcal/mol), the Hawkins GB/SA score (kcal/mol), the AMBER score, COSMO surface area of guest molecules (Å^2), COSMO volume of guest molecules (Å^3), COSMO volume in the cavity of the guest molecules (Å^3), Contact Score, number of the flexible torsion angles of guest molecules (pieces).

No.	Grid-Based score, kcal/mol	Hawkins GB/SA score, kcal/mol	AMBER score (GB)	A_{guest}, Å^2	V_{guest}, Å^3	$V_{in\ cavity}$, Å^3	Contact score	$N_{flexible\ torsions}$, pcs.
A1	–19.892433	–10.047655	–850.703491	121.73	123.63	115.93	–54	4
A2	–28.658304	–18.801785	–864.947144	139.56	155.88	127.68	–106	6
A3	–38.030979	–26.083462	–910.031311	165.65	202.48	165.27	–141	8
A4	–41.822193	–28.203060	–901.115479	171.99	225.29	156.39	–157	9
A5	–41.236294	–30.262899	–907.445679	206.80	246.20	167.61	–133	10
A6	–43.384472	–27.352457	–898.781555	275.70	362.24	163.38	–119	15
A7	–38.027222	–23.561420	–912.183533	174.02	208.80	141.38	–132	8
A8	–42.205219	–30.962749	–882.420715	159.23	212.81	151.08	–153	9
A9	–33.982277	–25.512693	no data	175.84	214.66	136.04	–98	8
A10	–33.803642	–26.868803	–898.072998	189.13	216.19	131.35	–112	5
A11	–38.198441	–28.619743	–889.687622	211.08	248.84	178.18	–116	7
A12	–35.505238	–24.106033	–868.904785	170.25	204.53	129.42	–113	8
A13	–38.890987	–27.765104	–867.632141	175.75	217.16	127.42	–130	9
A14	–40.595619	–27.018045	–878.344543	217.32	271.57	126.46	–131	11
A15	–39.981243	–27.541199	–886.495056	178.36	212.92	131.21	–121	3
A16	–25.289370	–15.852461	–741.131836	138.59	141.88	115.18	–66	4
A17	–35.626614	–22.603132	–781.583130	194.50	228.33	157.61	–87	8
A18	–31.963324	–21.967766	–811.640686	156.33	174.56	125.38	–82	6
A19	–37.412491	–25.646645	–805.641785	187.56	221.60	158.82	–110	8
A20	–35.380684	–22.617039	–797.355774	185.61	227.50	137.90	–95	8
A21	–40.361664	–25.852133	–782.911316	226.54	294.41	144.24	–142	11
A22	–34.928528	–20.690693	–782.609314	189.15	240.03	140.97	–134	9
A23	–41.450523	–25.142977	–797.080688	220.65	287.02	141.63	–116	11
A24	–46.972740	–37.574192	–896.076416	377.54	452.17	150.64	–147	13
A25	–45.338470	–33.399345	–953.485291	278.03	319.73	186.40	–111	9
A26	–31.482944	–27.112339	–927.747070	269.07	301.15	144.59	–106	8
A27	–44.992348	–32.207470	–973.252014	276.28	311.19	174.50	–125	10
A28	–47.545422	–34.946468	–919.839050	311.73	358.32	167.23	–123	12

Table S3. Properties of the molecules used to test the stability constant prediction (test set). The Grid-Based score (kcal/mol), the Hawkins GB/SA score (kcal/mol), the AMBER score, COSMO surface area of guest molecules (Å^2), Contact Score, number of the flexible torsion angles of guest molecules (pieces).

No.	Grid-Based score, kcal/mol	Hawkins GB/SA score, kcal/mol	AMBER score (GB)	A_{guest}, Å^2	Contact score	$N_{flexible\ torsions}$, pcs.
29	–23.847055	–16.496090	–887.351990	137.02	–107	3
30	–31.464674	–17.498480	–883.487671	232.66	–137	2
31	–36.894176	–28.954515	–863.386414	310.90	–107	8
32	–42.191219	–30.727964	–903.000488	350.10	–131	4
33	–37.905037	–30.234428	–893.212952	293.22	–113	5
34	–24.417015	–18.684515	–902.493896	299.98	–73	1
35	–35.492619	–27.145254	–902.371887	295.33	–130	0
36	–26.305637	–20.024298	–896.177063	316.72	–123	4
37	–31.125675	–20.477175	–894.306335	187.65	–112	0
38	–34.643444	–24.647732	–896.276855	218.47	–111	1
39	–38.533604	–22.595064	–897.334595	230.38	–117	2
40	–37.965252	–24.352243	–900.885193	227.67	–117	2
41	–39.633873	–26.627748	–913.838501	224.55	–119	2
42	–38.761040	–25.520107	–914.569946	230.93	–107	2
43	–34.711761	–24.206516	–903.833252	302.17	–91	6
44	–33.385254	–20.507299	–895.850037	300.69	–105	6
45	–40.358395	–28.555122	–919.821228	304.82	–114	6
46	–34.171917	–25.602039	–870.146423	226.11	–118	2
47	–32.750591	–25.139807	–857.248474	224.08	–120	2
48	–34.402527	–27.175968	–862.219482	220.40	–124	2
49	–34.228607	–26.587620	–885.076050	227.18	–106	2
50	–34.641987	–26.084553	–903.302124	300.16	–104	6
51	–31.503916	–22.299299	–882.973816	299.01	–104	6
52	–31.643497	–16.227791	–905.259460	144.45	–57	2

Table S4. The identification of the corresponding regression groups with linear equations. Molecules no. 29–52 were docked with the Hawkins GB/SA score and the Hawkins GB/SA score was estimated too from the equations I.–III. (article Table). The smallest absolute error (colored gray) between the "docked" and "fitted, estimated" Hawkins GB/SA score predicted the corresponding regression group (I., II., or III.).

	Hawkins GB/SA Score function/(kcal/mol)							
No.	Docking	I. regression	a. err.	II. regression	a. err.	III. regression	a. err.	Group
29	–16.50	–21.38	4.89	–14.74	1.76	–18.19	1.69	III.
30	–17.50	–31.37	13.88	–24.83	7.34	–30.81	13.31	II.
31	–28.95	–31.39	2.43	–24.93	4.03	–34.56	5.61	I.
32	–30.73	–39.88	9.15	–35.55	4.82	–42.51	11.79	II.
33	–30.23	–33.74	3.50	–29.40	0.83	–34.97	4.74	II.
34	–18.68	–28.73	10.05	–17.56	1.12	–28.32	9.64	II.
35	–27.15	–37.31	10.16	–31.84	4.69	–36.99	9.84	II.
36	–20.02	–30.27	10.25	–16.39	3.63	–33.37	13.34	II.
37	–20.48	–29.71	9.24	–26.73	6.25	–25.96	5.48	III.
38	–24.65	–31.95	7.30	–29.87	5.23	–29.21	4.56	III.
39	–22.60	–33.86	11.26	–33.46	10.87	–31.75	9.16	III.
40	–24.35	–33.46	9.11	–32.78	8.42	–31.34	6.98	III.
41	–26.63	–34.24	7.61	–34.79	8.16	–31.90	5.27	III.
42	–25.52	–33.55	8.03	–33.77	8.25	–31.09	5.57	III.
43	–24.21	–30.79	6.58	–24.52	0.32	–32.40	8.19	II.
44	–20.51	–30.69	10.19	–22.87	2.36	–32.89	12.38	II.
45	–28.56	–34.71	6.15	–31.27	2.71	–36.56	8.00	II.
46	–25.60	–31.57	5.97	–28.19	2.58	–29.86	4.25	II.
47	–25.14	–30.88	5.74	–26.46	1.32	–29.32	4.18	II.
48	–27.18	–31.71	4.53	–28.45	1.27	–30.00	2.82	II.
49	–26.59	–31.11	4.52	–28.30	1.71	–29.01	2.42	II.
50	–26.08	–31.25	5.16	–24.39	1.69	–33.25	7.16	II.
51	–22.30	–29.65	7.35	–20.60	1.70	–31.97	9.67	II.
52	–16.23	–24.08	7.86	–25.43	9.20	–18.05	1.82	III.

Table S5. The predicted stability constants. From equations I.–III. (article Figure 4), the log Ks were calculated using the "docked" Hawkins GB/SA scores. The calculated best log Ks with the smallest absolute error are colored gray. The theoretical average absolute error from these docking methods is ±0.49 log K unit. From the predicted regression groups based on (article Table) we got the average absolute error ± 0.92 log K units for the predicted stability constants.

No.	Log K from I.	Err.	Log K from II.	Err.	Log K from III.	Err.	Calculated best log K	Err.	Predicted group	Predicted log K	Err.
29	0.98	3.69	4.07	0.61	6.10	1.43	4.07	0.61	III.	6.10	1.43
30	1.26	2.66	4.28	0.36	6.32	2.40	4.28	0.36	II.	4.28	0.36
31	4.42	0.12	6.70	2.40	8.73	4.43	4.42	0.12	I.	4.42	0.12
32	4.91	0.46	7.07	2.62	9.11	4.66	4.91	0.46	II.	7.07	2.62
33	4.78	3.25	6.97	1.05	9.00	0.98	9.00	0.98	II.	6.97	1.05
34	1.59	3.07	4.53	0.13	6.57	1.91	4.53	0.13	II.	4.53	0.13
35	3.92	0.35	6.31	2.04	8.35	4.08	3.92	0.35	II.	6.31	2.04
36	1.96	3.02	4.81	0.17	6.85	1.87	4.81	0.17	II.	4.81	0.17
37	2.08	2.87	4.91	0.05	6.94	1.99	4.91	0.05	III.	6.94	1.99
38	3.23	2.67	5.79	0.12	7.82	1.92	5.79	0.12	III.	7.82	1.92
39	2.67	4.56	5.35	1.88	7.39	0.16	7.39	0.16	III.	7.39	0.16
40	3.15	4.02	5.72	1.44	7.76	0.60	7.76	0.60	III.	7.76	0.60
41	3.78	3.30	6.20	0.88	8.24	1.16	6.20	0.88	III.	8.24	1.16
42	3.47	3.66	5.97	1.16	8.01	0.88	8.01	0.88	III.	8.01	0.88
43	3.11	3.38	5.69	0.80	7.73	1.24	5.69	0.80	II.	5.69	0.80
44	2.09	4.57	4.91	1.75	6.95	0.29	6.95	0.29	II.	4.91	1.75
45	4.31	2.79	6.61	0.49	8.65	1.55	6.61	0.49	II.	6.61	0.49
46	3.50	2.87	5.99	0.38	8.03	1.66	5.99	0.38	II.	5.99	0.38
47	3.37	3.16	5.89	0.64	7.93	1.40	5.89	0.64	II.	5.89	0.64
48	3.93	2.80	6.32	0.41	8.36	1.62	6.32	0.41	II.	6.32	0.41
49	3.77	2.72	6.20	0.29	8.23	1.75	6.20	0.29	II.	6.20	0.29
50	3.63	1.99	6.09	0.47	8.13	2.50	6.09	0.47	II.	6.09	0.47
51	2.59	3.72	5.29	1.01	7.33	1.03	5.29	1.01	II.	5.29	1.01
52	0.91	4.00	4.01	0.90	6.05	1.14	4.01	0.90	III.	6.05	1.14
							Average abs. err.:	0.48		Average abs. err.:	0.92