Ligand binding constants of the cucurbit[7]uril predicted with molecular docking: A Theoretical Study; Turkish Journal of Chemistry; (2017)

Balázs ROÓSZ, Tamás KÖRTVÉLYESI†, Béla VISKOLCZ

Supplemental statistical data

1. Statistical data of the fitted equations using RCommander software.[1,2]

1.1. 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

Residuals:

Min., 1Q, Median, 3Q, Max.

-2.14751, -0.89746, 0.05731, 0.78542, 2.36475

Coefficients, Estimate Std. Error, t value, Pr(>|t|)

(Intercept) -1.640e+01, 5.898e+00, -2.781, 0.03196, *

Grid Based -1.539e-04, 3.276e-05, -4.698, 0.00333, **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.692 on 6 degrees of freedom

Multiple R-squared: 0.7863, Adjusted R-squared: 0.7506

F-statistic: 22.07 on 1 and 6 DF, p-value: 0.003333

Analysis of Variance Table

Response: lnK

Sum Sq., Df, F value, Pr(>F)

(Intercept) 22.146, 1, 7.7329, 0.031963, *

Grid Based 63.210, 1, 22.0713, 0.003333, **

Residuals 17.183, 6

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

II., ●: ln K = - (1.0987 ± 0.2146) × 10^{-4} × [Grid Score] - (3.9675 ± 3.4323); R^2 = 0.7238

Residuals:

Min., 1Q, Median, 3Q, Max.

-1.4221, -0.5291, -0.1784, 0.4746, 1.6604

Coefficients, Estimate Std. Error, t value, Pr(>|t|)

(Intercept) -3.968e+00, 3.432e+00, -1.156, 0.274573,

Grid Based -1.099e-04, 2.146e-05, -5.119, 0.000452, ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9645 on 10 degrees of freedom

Multiple R-squared: 0.7238, Adjusted R-squared: 0.6961

F-statistic: 26.2 on 1 and 10 DF, p-value: 0.0004517

Analysis of Variance Table

Response: lnK

Sum Sq., Df, F value, Pr(>F)

(Intercept) 1.2430, 1, 1.3362, 0.2745729,

Grid Based 24.3715, 1, 26.1993, 0.0004517, ***

Residuals 9.3023, 10

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

III., ▲: ln K = - (1.0646 ± 0.1937) × 10^{-4} × [Grid Score] + (2.3953 ± 2.6188); R^2 = 0.8342)

Residuals:

Min., 1Q, Median, 3Q, Max

-1.8098, -1.1082, 0.4305, 0.8222, 1.8846

Coefficients, Estimate Std. Error, t value, Pr(>|t|)

(Intercept) 2.395e+00, 2.619e+00, 0.915, 0.39563,

Grid Based -1.065e-04, 1.938e-05, -5.494, 0.00152, **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.451 on 6 degrees of freedom

Multiple R-squared: 0.8342, Adjusted R-squared: 0.8066

F-statistic: 30.19 on 1 and 6 DF, p-value: 0.001523

Analysis of Variance Table

Response: lnK

Sum Sq., Df, F value, Pr(>F)

(Intercept) 1.762, 1, 0.8366, 0.395634,

Grid Based 63.586, 1, 30.1884, 0.001523, **

Residuals 12.638, 6

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

1.2.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

Residuals:

Min. 1Q, Median, 3Q, Max.

-2.08983, -0.81339, -0.07719, 0.74697, 1.80568

Coefficients, Estimate Std. Error, t value, Pr(>|t|)

(Intercept) -8.218e+00, 2.884e+00, -2.850, 0.024698, *

Hawkins GB/SA -1.519e-04, 2.252e-05, -6.745, 0.000266, ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.248 on 7 degrees of freedom

Multiple R-squared: 0.8667, Adjusted R-squared: 0.8476

F-statistic: 45.5 on 1 and 7 DF, p-value: 0.0002662

Analysis of Variance Table

Response: lnK

Sum Sq., Df, F value, Pr(>F)

(Intercept) 12.639, 1, 8.121, 0.0246976, *

Hawkins.GB/SA 70.807, 1, 45.497, 0.0002662, ***

Residuals 10.894, 7

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

II., ●: ln K = - (1.1625 ± 0.1611) × 10^{-4} × [Hawkins GB/SA Score] + (1.3382 ± 1.7457); R^2 = 0.8525

Residuals:

Min., 1Q, Median, 3Q, Max

-0.85059, -0.50183, 0.09695, 0.37300, 0.75215

Coefficients:

Estimate Std. Error, t value, Pr(>|t|)

(Intercept) 1.338e+00, 1.746e+00, 0.767, 0.463

Hawkins GB/SA -1.163e-04, 1.612e-05, -7.213, 5.01e-05, ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5669 on 9 degrees of freedom

Multiple R-squared: 0.8525, Adjusted R-squared: 0.8361

F-statistic: 52.02 on 1 and 9 DF, p-value: 5.014e-05

Analysis of Variance Table

Response: lnK

Sum Sq., Df, F value, Pr(>F)

(Intercept) 0.1888, 1, 0.5876, 0.463,

Hawkins.GB/SA 16.7185, 1, 52.0229, 5.014e-05, ***

Residuals 2.8923, 9

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

III., ▲: ln K = - (1.1610 ± 0.1659) × 10^{-4} × [Hawkins GB/SA Score] + (6.0427 ± 1.5521); R^2 = 0.8908

Residuals:

Min., 1Q, Median, 3Q, Max.

-1.79119, -0.66762, 0.06449, 0.79399, 1.56444

Coefficients, Estimate Std. Error, t value, Pr(>|t|)

(Intercept) 6.043e+00, 1.552e+00, 3.893, 0.008046, **

Hawkins GB/SA -1.161e-04, 1.659e-05, -6.997, 0.000424, ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.178 on 6 degrees of freedom

Multiple R-squared: 0.8908, Adjusted R-squared: 0.8726

F-statistic: 48.96 on 1 and 6 DF, p-value: 0.0004244

Analysis of Variance Table

Response: lnK

Sum Sq., Df, F value, Pr(>F)

(Intercept) 21.022, 1, 15.158, 0.0080458, **

Hawkins GB/SA 67.903, 1, 48.962, 0.0004244, ***

Residuals 8.321, 6

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

1.3.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.

A: □; ln K = - (1.0260 ± 0.2660) × 10^{-1} × [Amber Score] - (86.6670 ± 25.0644); R^2 = 0.8814

Residuals:

1, 2, 3, 4

-0.9853, -0.2480, 1.4300, -0.1967

Coefficients, Estimate Std. Error, t value, Pr(>|t|)

(Intercept) -86.66703, 25.06445, -3.458, 0.0744, .

AMBER score -0.10260, 0.02661, -3.856, 0.0612, .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.248 on 2 degrees of freedom

Multiple R-squared: 0.8814, Adjusted R-squared: 0.8222

F-statistic: 14.87 on 1 and 2 DF, p-value: 0.06115

Analysis of Variance Table

Response: ln.K.

Sum Sq., Df, F value, Pr(>F)

(Intercept) 18.6265, 1, 11.956, 0.07442, .

AMBER score 23.1635, 1, 14.868, 0.06115, .

Residuals 3.1158, 2

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

B: ◇; ln K = - (7.9810 ± 3.3674) × 10^{-2} × [Amber Score] - (59.6434 ± 30.4033); R^2 = 0.7374

Residuals:

1, 2, 3, 4

0.4723, -1.1205, 0.5068, 0.1415

Coefficients, Estimate Std. Error, t value, Pr(>|t|)

(Intercept) -59.64341, 30.40328, -1.962, 0.189,

AMBER score -0.07981, 0.03367, -2.370, 0.141,

Residual standard error: 0.9369 on 2 degrees of freedom

Multiple R-squared: 0.7374, Adjusted R-squared: 0.6062

F-statistic: 5.617 on 1 and 2 DF, p-value: 0.1413

Analysis of Variance Table

Response: ln.K

Sum Sq., Df, F value, Pr(>F)

(Intercept) 3.3780, 1, 3.8484, 0.1888

AMBER score 4.9307, 1, 5.6174, 0.1413

Residuals 1.7555 2

C: ◯; ln K = - (7.6047 ± 1.1759) × 10^{-2} × [Amber Score] - (52.4759 ± 10.3430); R^2 = 0.8745

Residuals:

Min., 1Q, Median, 3Q, Max.

-0.6902, -0.5532, 0.1002, 0.2969, 0.8986

Coefficients, Estimate Std. Error, t value, Pr(>|t|)

(Intercept) -52.47595, 10.34299, -5.074, 0.002280, **

AMBER score -0.07605, 0.01176, -6.467, 0.000648, ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6243 on 6 degrees of freedom

Multiple R-squared: 0.8745, Adjusted R-squared: 0.8536

F-statistic: 41.83 on 1 and 6 DF, p-value: 0.0006485

Analysis of Variance Table

Response: ln.K

Sum Sq., Df, F value, Pr(>F)

(Intercept) 10.0311, 1, 25.741, 0.0022799, **

AMBER score 16.2988, 1, 41.825, 0.0006485, ***

Residuals 2.3381, 6

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

D: △; ln K = - (6.8172 ± 0.9277) × 10^{-2} × [Amber Score] - (40.7651 ± 7.6928); R^2 = 0.8852

Residuals:

Min., 1Q, Median, 3Q, Max

-1.81734, -0.81203, -0.05545, 1.18057, 1.67760

Coefficients, Estimate Std. Error, t value, Pr(>|t|)

(Intercept) -40.765086, 7.692761, -5.299, 0.001124, **

AMBER score -0.068172, 0.009277, -7.348, 0.000156, ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.275 on 7 degrees of freedom

Multiple R-squared: 0.8852, Adjusted R-squared: 0.8688

F-statistic: 54 on 1 and 7 DF, p-value: 0.0001561

Analysis of Variance Table

Response: ln K

Sum Sq., Df, F value, Pr(>F)

(Intercept) 45.662, 1, 28.081, 0.0011244, **

AMBER score 87.806, 1, 53.999, 0.0001561, ***

Residuals 11.382, 7

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

2. McQSAR 2D-QSAR study[3]

2.1. We performed a short QSAR study to find other molecular descriptors that are increasing the number of displaced cavity water molecules.

From the intersections of the lnK – Hawkins GB/SA Score equations the number of displaced cavity waters (nwater) are estimated. In this case an approximation used, that the different “nwater” number causes the difference between I., II., III. molecular groups found between the lnK – Hawkins GB/SA Score intersections and lnK – Grid-Based Score intersections (an average slope was applied to calculate the intersection for every data point). We tried to find a linear equation using the McQSAR software

(genetic algorithm) with PaDEL 2D molecular descriptors.[3,4] We found that increasing the number of hydrogen bond donors (OH, NH, etc.), the partial negative surface area (sum of surface area on negative parts of molecule), the mean atomic (Sanderson) electronegativities, the Lipoaffinity index or the length over breadth ratio for the rotation that results in the minimum area, all increase the number of displaced cavity waters (nwater).

If a good connection could be find between molecular descriptors and the number of replaced cavity waters for a specific target molecule cavity and homologous ligand molecules, it would make easier to predict the ligand binding stability constants for these systems with a fast and computationally very cheap method combined with molecular docking calculations.

2.2.The best equations found with the genetic algorithm -GA- search (McQSAR).[3]

I. From 63 selected, active molecular descriptors (collinearity cutoff < 0.5), generating 200000 GA generation.

nwater = -1.95175 - 2.76823 × AATS3p + 3.54641 × hmin + 2.48137 × SpMax4_Bhs

R2:0.862673, LOF:0.206031, **score:0.827234**, rel_E:1.79769e+308, leverage:28 1.61129 -0.917059 -0.603291 4.29524 0.301728 0.230158, validation-R2_0: 0.999572

AATS3p: Average Broto-Moreau autocorrelation - lag 3 / weighted by polarizabilities (Autocorrelation Descriptor)

hmin: Minimum H E-State (Electrotopological State Atom Type Descriptor)

SpMax4_Bhs: Largest absolute eigenvalue of Burden modified matrix - n 4 / weighted by relative I-state (Burden Modified Eigenvalues Descriptor)

nwater = 1.79354 × ETA_Eta_B + 1.81519 × SpMax4_Bhs - 19.1064 + 0.0958566 × AATS3i + 2.72916 × hmin

R2:0.907951, LOF:0.168704, **score:0.870145**, rel_E:1.79769e+308, leverage:28 1.4564 -0.11945 -0.000684695 -0.543022 0.424754 -0.00783148 0.0661079 0.000151814 -0.000958617 3.98549, validation-R2_0: 0.943468

ETA_Eta_B: Branching index EtaB (Extended Topochemical Atom Descriptor)

AATS3i: Average Broto-Moreau autocorrelation - lag 3 / weighted by first ionization potential (Autocorrelation Descriptor)

II. From 32 selected, relevant, active molecular descriptors (collinearity cutoff < 0.5), generating 200000 GA generation.

nwater = 0.110947 × Kier3 + 0.301845 × nHBDon_Lipinski + 0.0113936 × PNSA-1

R2:0.782151, LOF:0.297799, **score:0.747081**, rel_E:1.79769e+308, leverage:28 0.00331435 0.000478959 -0.000137373 0.020998 -0.000187991 8.8773e-06

nwater = 0.0240235 × VE3_DzZ + 0.133058 × Kier3 + 0.00928624 × PNSA-1 + 0.348894 × nHBDon_Lipinski

R2:0.824585, LOF:0.290149, **score:0.787652**, rel_E:1.79769e+308, leverage:28 0.000575578 -0.000529744 5.04904e-05 -0.00112724 0.00380191 -0.000183843

0.00151644 1.33064e-05 -0.000286874 0.0232056, validation-R2_0: 0.946345

Kier3: Third kappa (κ) shape index (Kappa Shape Indices Descriptor)

nHBDon_Lipinski: Number of hydrogen bond donors, using Lipinski's definition: Any OH or NH. Each available hydrogen atom is counted as one hydrogen bond donor. (PaDEL H-Bond Donor Count Descriptor)

PNSA-1:Partial negative surface area -- sum of surface area on negative parts of molecule (CPSA Descriptor)

VE3_DzZ: Logarithmic coefficient sum of the last eigenvector from Barysz matrix / weighted by atomic number (Barysz Matrix Descriptor)

III. From 24 selected, most relevant, active molecular descriptors (collinearity cutoff < 0.5), generating 100000 GA generation.

nwater = 44.3754 × Mse -42.9202 + 0.289321 × LipoaffinityIndex + 0.356745 × nHBDon_Lipinski

R2:0.789046, LOF:0.316492, **score:0.757159**, rel_E:1.79769e+308, leverage:28 0.345087 -0.0495708 -0.0511585 0.00858002 0.00533594 0.0224087, validation-R2_0: 0.911133

Mse:Mean atomic Sanderson electronegativities, scaled on carbon atom (Constitutional Descriptor)

LipoaffinityIndex: Lipoaffinity index (Electrotopological State Atom Type Descriptor)

nwater = -33.2138 + 0.318628 × nHBDon_Lipinski + 33.4631 × Mse + 0.259696 × LipoaffinityIndex + 0.00884946 × PNSA-1

R2:0.866674, LOF:0.244356, **score:0.786638**, rel_E:1.79769e+308, leverage:28

0.0241567 -0.0169621 0.00496521 -0.000247681 1.01408 -0.0568236 -0.00484545 0.00865865 5.25311e-05 3.50951e-05

IV. From 16 strongly selected, most relevant, active molecular descriptors (collinearity cutoff < 0.5), generating 50000 GA generation.

nwater = 0.41089 × nHBDon_Lipinski + 0.0082495 × PNSA-1 + 0.597001 × LOBMIN

R2:0.773708, LOF:0.309342, **score:0.727001**, rel_E:1.79769e+308, leverage:28

0.0253803 -0.0004437 0.0212472 2.02414e-05 -0.00131526 0.101414

LOBMIN: The L/B ratio for the rotation that results in the minimum area (Length Over Breadth Descriptor)

nwater = 0.0102204 × PNSA-1 + 0.618517 × LOBMIN + 0.375758 × nHBDon_Lipinski - 0.814903 × RNCG

R2:0.797969, LOF:0.334173, **score:0.734917**, rel_E:1.79769e+308, leverage:28

2.7017e-05 -0.00124129 -0.000564479 -0.00280153 0.102221 0.0199287 -0.0305849 0.0275332 0.0499386 1.15835

RNCG: Relative negative charge -- most negative charge / total negative charge (CPSA Descriptor)

Used software in this supplemental material.:

RCommander[1,2] (under Linux use package manager to install)

McQSAR[3] (download and install manually)

PaDEL-Descriptor[4] (download and install manually)

All calculations were done under Ubuntu Linux operating system.

References of the Supplemental statistical data

1. Fox, J. *J. Stat. Software, Artic.* **2005**, *14*, 1–42.
2. Fox, J. *Using the R Commander: A Point-and-Click Interface for R*; Chambers, J. M., Hothorn, T., Lang, D. T., Wickham, H., Eds.; Chapman and Hall/CRC: Abingdon, **2016**.
3. Vainio, M. J.; Johnson, M. S. *J. Chem. Inf. Model.* **2005**, *45*, 1953–1961.
4. Yap, C. W. *J. Comput. Chem.* **2011**, *32*, 1466–1474.