**SUPPORTING INFORMATION:**

**Taba: A Tool to Analyze the Binding Affinity**

**Amauri Duarte da Silva, ,2 Gabriela Bitencourt-Ferreira,1 and Walter Filgueira de Azevedo Junior1, ,\***

**Correspondence to: Walter Filgueira de Azevedo Junior (E-mail:** [**walter@azevedolab.net**](mailto:walter@azevedolab.net)**)**

*Laboratory of Computational Systems Biology, Pontifical Catholic University of Rio Grande do Sul (PUCRS), Ipiranga Avenue, 6681 Partenon - Porto Alegre/RS. Brazil. Zip Code: 90619-900.*

*Specialization Program in Bioinformatics, Pontifical Catholic University of Rio Grande do Sul (PUCRS), Ipiranga Avenue, 6681 Partenon - Porto Alegre/RS. Brazil. Zip Code: 90619-900*

**Supplementary material 1.** PDB access codes for all structures in the CDKKi dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PDB access code | Ligand Identification | Ligand Chain | Ligand Number | Ki (nM) | Test Set |
| 1E1V | CMG | A | 401 | 12000 | 1 |
| 1E1X | NW1 | A | 401 | 1300 | 0 |
| 1H1S | 4SP | A | 1298 | 6 | 0 |
| 1JSV | U55 | A | 400 | 2000 | 1 |
| 1OGU | ST8 | A | 1298 | 2400 | 0 |
| 1PXM | CK5 | A | 500 | 60 | 1 |
| 1PXN | CK6 | A | 500 | 195 | 0 |
| 1PXO | CK7 | A | 500 | 2 | 1 |
| 1PXP | CK8 | A | 500 | 220 | 0 |
| 1PYE | PM1 | A | 700 | 386 | 1 |
| 1V1K | 3FP | A | 299 | 35000 | 1 |
| 2CLX | F18 | A | 1299 | 13300 | 0 |
| 2EXM | ZIP | A | 400 | 78000 | 0 |
| 2FVD | LIA | A | 299 | 3 | 0 |
| 2XMY | CDK | A | 500 | 0.11 | 1 |
| 2XNB | Y8L | A | 1299 | 149 | 1 |
| 3BLR | CPB | A | 940 | 3 | 0 |
| 3DDQ | RRC | A | 299 | 250 | 0 |
| 3LFN | A27 | A | 299 | 3160 | 0 |
| 3LFS | A07 | A | 299 | 2500 | 1 |
| 3MY5 | RFZ | A | 300 | 65000 | 0 |
| 4ACM | 7YG | A | 1302 | 210 | 0 |
| 4BCK | T3E | A | 1298 | 4 | 0 |
| 4BCM | T7Z | A | 1297 | 123 | 0 |
| 4BCN | T9N | A | 1299 | 12 | 0 |
| 4BCO | T6Q | A | 1299 | 131 | 0 |
| 4BCP | T3C | A | 1299 | 568 | 0 |
| 4BCQ | TJF | A | 1296 | 147 | 0 |
| 4EOP | 1RO | A | 301 | 890 | 0 |
| 4NJ3 | 2KD | A | 301 | 140 | 0 |
| 5D1J | 56H | A | 4000 | 38 | 0 |

**Supplementary material 2.** Predictive performance of scoring functions (training set).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scoring Functions | ρ | p-value1 | R | p-value2 |
| Free Energya | 0.358 | 0.1018 | 0.000 | 0.9506 |
| Final Intermolecular Energya | 0.244 | 0.2731 | 0.000 | 0.9501 |
| vdW+Hbond+desolv Energya | 0.248 | 0.2659 | 0.000 | 0.9501 |
| Electrostatic Energya | 0.015 | 0.9462 | 0.000 | 0.9891 |
| Final Total Internal Energya | 0.407 | 0.0600 | 0.143 | 0.0824 |
| Torsional Free Energya | -0.24 | 0.2827 | 0.093 | 0.1665 |
| Plants Scoreb | 0.351 | 0.1095 | 0.158 | 0.0665 |
| MolDock Scoreb | 0.299 | 0.1759 | 0.167 | 0.0590 |
| Rerank Scoreb | 0.278 | 0.2095 | 0.214 | 0.0302 |
| Interaction Scoreb | 0.274 | 0.2164 | 0.159 | 0.0661 |
| Protein Scoreb | 0.329 | 0.1353 | 0.176 | 0.0517 |
| Water Scoreb | -0.101 | 0.6561 | 0.034 | 0.4097 |
| Internal Scoreb | -0.097 | 0.6690 | 0.032 | 0.4253 |
| Electrostatic Scoreb | -0.052 | 0.8196 | 0.001 | 0.8702 |
| Electrostatic Long Scoreb | -0.257 | 0.2489 | 0.056 | 0.2883 |
| H-Bond Scoreb | 0.188 | 0.4012 | 0.028 | 0.4578 |
| Ligand Efficiency 1 Scoreb | -0.120 | 0.5956 | 0.002 | 0.8365 |
| Ligand Efficiency 3 Scoreb | 0.034 | 0.8810 | 0.050 | 0.3177 |
| Affinity Scorec | 0.267 | 0.2304 | 0.015 | 0.5910 |
| Gauss1 Scorec | -0.249 | 0.2648 | 0.121 | 0.1124 |
| Gauss2 Scorec | -0.294 | 0.1846 | 0.152 | 0.0732 |
| Repulsion Scorec | -0.018 | 0.9364 | 0.000 | 0.9974 |
| Hydrophobic Scorec | -0.132 | 0.5577 | 0.002 | 0.8612 |
| Hydrogen Scorec | -0.018 | 0.9384 | 0.000 | 0.9289 |
| Taba (3 variables, d≤ 4.5 Å) | 0.558 | 0.0063 | 0.576 | 0.0086 |

aAutoDock 4, bMolegro Virtual Docker (MVD), cAutoDock Vina. P-value1 and p-value2 are related to ρ and R, respectively.

**Supplementary material 3A.** Predictive performance of Taba scoring functions (training set).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Taba Scoring Functions | Radius (Å) | Number of Variables | Regression Method | Standard deviation | ρ | p-value1 | R | p-value2 |
| α0 = - 6.596172; αC,F = - 0.438606; αO,O = - 0.128109  dC,F = 3.2166; dO,O = 3.0478 | 3.5 | 2 | Linear Regression | 1.28225 | 0.565 | 0.0076 | 0.353 | 0.1166 |
| α0= - 6.588944; αC,F = - 0.453532; αN,O = - 0.368455; αO,O = - 0.062283  dC.F = 3.2166; dN.O = 3.2153; dO.O = 3.0478 | 3.5 | 3 | Linear Regression | 1.25444 | 0.625 | 0.0027 | 0.457 | 0.0374 |
| α0= - 6.582380; αC.C= - 0.223546; αC.F= - 0.480265; αN.O= - 0.392575 αO.O= - 0.096632  dC.C = 3.3443; dC.F = 3.2166; dN.O = 3.2153; dO.O = 3.0478 | 3.5 | 4 | Linear Regression | 1.26774 | 0.677 | 0.0008 | 0.489 | 0.0244 |
| α0= - 6.609189; αC.S= - 0.240171; αC.F= - 0.435923  dC.S = 4.1193; dC.F = 3.8503 | 4.5 | 2 | Linear Regression | 1.26305 | 0.541 | 0.0113 | 0.388 | 0.0822 |
| α0= - 6.581356; αC.N= - 0.111232; αC.O= - 0.406456; αN.F= - 0.353717  dC.N = 3.9946; dC.O = 3.8862; dN.F= 4.2167 | 4.5 | 3 | Elastic Net CV | 1.16488 | 0.558 | 0.0086 | 0.576 | 0.0063 |
| α0= - 6.585354; αC.F= - 0.348295; αN.N= - 0.030841; αN.O =- 0.226548; αO.S = - 0.308260  dC.F = 3.8503; dN.N = 3.7631; dN.O = 3.8262; dO.S = 3.9396 | 4.5 | 4 | Elastic Net CV | 1.24219 | 0.572 | 0.0068 | 0.537 | 0.0120 |
| α0= - 6.596825; αN.N = 0.054457; αO.S = - 0.473969  dN.N = 5.0996; dO.S = 5.2411 | 6 | 2 | Linear Regression | 1.28430 | 0.618 | 0.0028 | 0.349 | 0.1211 |
| α0= - 6.598658; αC.N= - 2.249965; αN.N= 2.413916; αO.S= - 0.677687  dC.N = 5.0800; dN.N = 5.0996; dO.S = 5.2411 | 6 | 3 | Linear Regression | 1.22439 | 0.604 | 0.0037 | 0.496 | 0.0222 |
| α0= - 6.591343, αC.N = - 2.077791; αC.F = - 0.418423; αN.N = 2.107410; αO.S = - 0.588023  dC.N = 4.9532; dC.F = 4.6903; dN.N = 5.0996; dO.S = 5.2411 | 6 | 4 | Linear Regression | 1.16781 | 0.644 | 0.0016 | 0.595 | 0.0044 |
| α0= - 6.593903; αC.F = - 0.450222; αO.S = - 0.464469  dC.F = 5.7584; dO.S = 6.2435 | 7.5 | 2 | Linear Regression | 1.18399 | 0.575 | 0.0064 | 0.504 | 0.0200 |
| α0= - 6.601670; αC.S = - 0.188280; αC.F = - 0.219935; αN.S = - 0.070018  dC.S = 6.1533; dC.F = 5.7584; dN.S = 6.5248 | 7.5 | 3 | Ridge | 1.27537 | 0.583 | 0.0055 | 0.481 | 0.0271 |
| α0= - 6.625082; αC.S = - 1.232409; αC.F = - 0.466594; αN.N = 0.294947; αN.S = 0.608300  dC.S = 6.1533; dC.F = 5.7584; dN.N = 6.3197; dN.S = 6.5248 | 7.5 | 4 | Linear Regression | 1.23493 | 0.674 | 0.0008 | 0.527 | 0.0140 |
| α0= - 6.594082; αC.F= - 0.448794; αN.S = - 0.409427  dC.F = 6.9792; dN.S = 7.5769 | 9 | 2 | Linear Regression | 1.20849 | 0.584 | 0.0054 | 0.472 | 0.0309 |
| α0= - 6.631846; αC.S= - 2.022452; αN.N= 0.423672; αN.S= 1.278357  dC.S = 7.2947; dN.N = 7.4161; dN.S = 7.5769 | 9 | 3 | Linear Regression | 1.27558 | 0.604 | 0.0037 | 0.426 | 0.0540 |
| α0= - 6.624144; αC.N= 0.333580; αC.S= - 2.018330; αC.F= - 0.452698; αN.S= 1.307272  dC.N = 7.0712; dC.S = 7.2947; dC.F = 6.9792; dN.S = 7.5769 | 9 | 4 | Linear Regression | 1.20594 | 0.692 | 0.0005 | 0.558 | 0.0085 |

p-value1 and p-value2 are related to ρ and R, respectively. In the Taba scoring functions the atoms indicate which pair was included in the model. The distance (di,j) is in Å.

**Supplementary material 3B.** Predictive performance of Taba scoring functions (test set).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Taba Scoring Functions | Radius (Å) | Number of Variables | Regression Method | ρ | p-value1 | R | p-value2 |
| α0 = - 6.596172; αC,F = - 0.438606; αO,O = - 0.128109  dC,F = 3.2166; dO,O = 3.0478 | 3.5 | 2 | Linear Regression | 0.548 | 0.889 | -0.292 | 0.444 |
| α0= - 6.588944; αC,F = - 0.453532; αN,O = - 0.368455; αO,O = - 0.062283  dC.F = 3.2166; dN.O = 3.2153; dO.O = 3.0478 | 3.5 | 3 | Linear Regression | 0.109 | 0.781 | -0.356 | 0.348 |
| α0= - 6.582380; αC.C= - 0.223546; αC.F= - 0.480265; αN.O= - 0.392575 αO.O= - 0.096632  dC.C = 3.3443; dC.F = 3.2166; dN.O = 3.2153; dO.O = 3.0478 | 3.5 | 4 | Linear Regression | 0.350 | 0.356 | 0.352 | 0.354 |
| α0= - 6.609189; αC.S= - 0.240171; αC.F= - 0.435923  dC.S = 4.1193; dC.F = 3.8503 | 4.5 | 2 | Linear Regression | 0.176 | 0.651 | -0.387 | 0303 |
| α0= - 6.581356; αC.N= - 0.111232; αC.O= - 0.406456; αN.F= - 0.353717  dC.N = 3.9946; dC.O = 3.8862; dN.F= 4.2167 | 4.5 | 3 | Elastic Net CV | 0.783 | 0.013 | 0.794 | 0.011 |
| α0= - 6.585354; αC.F= - 0.348295; αN.N= - 0.030841; αN.O =- 0.226548; αO.S = - 0.308260  dC.F = 3.8503; dN.N = 3.7631; dN.O = 3.8262; dO.S = 3.9396 | 4.5 | 4 | Elastic Net CV | 0.067 | 0.845 | -0.290 | 0.449 |
| α0= - 6.596825; αN.N = 0.054457; αO.S = - 0.473969  dN.N = 5.0996; dO.S = 5.2411 | 6 | 2 | Linear Regression | 0.183 | 0.637 | 0.548 | 0.126 |
| α0= - 6.598658; αC.N= - 2.249965; αN.N= 2.413916; αO.S= - 0.677687  dC.N = 5.0800; dN.N = 5.0996; dO.S = 5.2411 | 6 | 3 | Linear Regression | 0.383 | 0.308 | 0.451 | 0.223 |
| α0= - 6.591343, αC.N = - 2.077791; αC.F = - 0.418423; αN.N = 2.107410; αO.S = - 0.588023  dC.N = 4.9532; dC.F = 4.6903; dN.N = 5.0996; dO.S = 5.2411 | 6 | 4 | Linear Regression | 0.250 | 0.516 | 0.402 | 0.283 |
| α0= - 6.593903; αC.F = - 0.450222; αO.S = - 0.464469  dC.F = 5.7584; dO.S = 6.2435 | 7.5 | 2 | Linear Regression | 0.109 | 0.781 | -0.288 | 0.452 |
| α0= - 6.601670; αC.S = - 0.188280; αC.F = - 0.219935; αN.S = - 0.070018  dC.S = 6.1533; dC.F = 5.7584; dN.S = 6.5248 | 7.5 | 3 | Ridge | 0.276 | 0.472 | 0.090 | 0.817 |
| α0= - 6.625082; αC.S = - 1.232409; αC.F = - 0.466594; αN.N = 0.294947; αN.S = 0.608300  dC.S = 6.1533; dC.F = 5.7584; dN.N = 6.3197; dN.S = 6.5248 | 7.5 | 4 | Linear Regression | 0.417 | 0.265 | 0.494 | 0.177 |
| α0= - 6.594082; αC.F= - 0.448794; αN.S = - 0.409427  dC.F = 6.9792; dN.S = 7.5769 | 9 | 2 | Linear Regression | 0.176 | 0.651 | -0.333 | 0.381 |
| α0= - 6.631846; αC.S= - 2.022452; αN.N= 0.423672; αN.S= 1.278357  dC.S = 7.2947; dN.N = 7.4161; dN.S = 7.5769 | 9 | 3 | Linear Regression | 0.917 | 0.001 | 0.813 | 0.008 |
| α0= - 6.624144; αC.N= 0.333580; αC.S= - 2.018330; αC.F= - 0.452698; αN.S= 1.307272  dC.N = 7.0712; dC.S = 7.2947; dC.F = 6.9792; dN.S = 7.5769 | 9 | 4 | Linear Regression | 0.300 | 0.433 | 0.622 | 0.074 |

p-value1 and p-value2 are related to ρ and R, respectively. In the Taba scoring functions the atoms indicate which pair was included in the model. The distance (di,j) is in Å.

**Supplementary material 4.** Predicted binding affinities for all scores avaiable in the program AutoDock 4 (all structures in CDKKi dataset).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PDB access code | Ligand Identification | Ligand Chain | Ligand Number | Ki (nM) | Test Set | Free Energy | Final Intermolecular Energy | vdW+Hbond+desolv Energy | Electrostatic Energy | Final Total Internal Energy | Torsional Free Energy |
| 1E1V | CMG | A | 401 | 12000 | 1 | 0.93 | -0.1 | -0.16 | 0.06 | -0.17 | 1.19 |
| 1E1X | NW1 | A | 401 | 1300 | 0 | -3.56 | -7.26 | -7.12 | -0.14 | 1.92 | 1.79 |
| 1H1S | 4SP | A | 1298 | 6 | 0 | -9.97 | -11.18 | -10.92 | -0.26 | -0.88 | 2.09 |
| 1JSV | U55 | A | 400 | 2000 | 1 | 374.26 | 372.95 | 372.68 | 0.27 | -0.18 | 1.49 |
| 1OGU | ST8 | A | 1298 | 2400 | 0 | -4.4 | -9.73 | -9.59 | -0.14 | 2.05 | 3.28 |
| 1PXM | CK5 | A | 500 | 60 | 1 | 2080 | 2080 | 2080 | 0.02 | -0.36 | 0.89 |
| 1PXN | CK6 | A | 500 | 195 | 0 | 2060 | 2060 | 2060 | -0.49 | -0.48 | 1.19 |
| 1PXO | CK7 | A | 500 | 2 | 1 | 21400 | 21400 | 21400 | -0.84 | -0.05 | 0.89 |
| 1PXP | CK8 | A | 500 | 220 | 0 | 17600 | 17600 | 17600 | -0.21 | -0.46 | 0.89 |
| 1PYE | PM1 | A | 700 | 386 | 1 | -8.19 | -8.29 | -8.15 | -0.14 | -1.98 | 2.09 |
| 1V1K | 3FP | A | 299 | 35000 | 1 | -5.49 | -6.82 | -7.22 | 0.4 | -1.95 | 3.28 |
| 2CLX | F18 | A | 1299 | 13300 | 0 | 18.44 | 12.99 | 12.83 | 0.16 | 3.66 | 1.79 |
| 2EXM | ZIP | A | 400 | 78000 | 0 | -4.81 | -5.39 | -5.31 | -0.08 | -0.31 | 0.89 |
| 2FVD | LIA | A | 299 | 3 | 0 | -9.92 | -10.31 | -9.94 | -0.37 | -1.39 | 1.79 |
| 2XMY | CDK | A | 500 | 0.11 | 1 | -8.05 | -10.07 | -9.87 | -0.2 | -0.66 | 2.68 |
| 2XNB | Y8L | A | 1299 | 149 | 1 | -8.47 | -9.43 | -9.84 | 0.41 | -0.54 | 1.49 |
| 3BLR | CPB | A | 940 | 3 | 0 | -9.67 | -9.61 | -9.66 | 0.05 | -1.85 | 1.79 |
| 3DDQ | RRC | A | 299 | 250 | 0 | -7.11 | -8.71 | -9.01 | 0.3 | -1.09 | 2.68 |
| 3LFN | A27 | A | 299 | 3160 | 0 | -11.6 | -11.97 | -11.85 | -0.12 | -1.42 | 1.79 |
| 3LFS | A07 | A | 299 | 2500 | 1 | -8.59 | -8.83 | -8.86 | 0.03 | -0.65 | 0.89 |
| 3MY5 | RFZ | A | 300 | 65000 | 0 | -5.69 | -5.69 | -5.73 | 0.04 | -0.6 | 0.6 |
| 4ACM | 7YG | A | 1302 | 210 | 0 | -10.92 | -11.39 | -11.34 | -0.06 | -2.21 | 2.68 |
| 4BCK | T3E | A | 1298 | 4 | 0 | -6.2 | -7.84 | -7.94 | 0.1 | -0.74 | 2.39 |
| 4BCM | T7Z | A | 1297 | 123 | 0 | -6.14 | -7.52 | -8 | 0.47 | -1.01 | 2.39 |
| 4BCN | T9N | A | 1299 | 12 | 0 | -5.06 | -7.57 | -8.04 | 0.47 | 0.42 | 2.09 |
| 4BCO | T6Q | A | 1299 | 131 | 0 | -7.33 | -9.18 | -10.03 | 0.85 | -0.23 | 2.09 |
| 4BCP | T3C | A | 1299 | 568 | 0 | -7.81 | -9.05 | -9.16 | 0.11 | -0.85 | 2.09 |
| 4BCQ | TJF | A | 1296 | 147 | 0 | -4.44 | -5.42 | -6.14 | 0.72 | -1.11 | 2.09 |
| 4EOP | 1RO | A | 301 | 890 | 0 | -3.18 | -10.31 | -10.23 | -0.08 | 5.64 | 1.49 |
| 4NJ3 | 2KD | A | 301 | 140 | 0 | -15.35 | -15.92 | -14.97 | -0.96 | -2.41 | 2.98 |
| 5D1J | 56H | A | 4000 | 38 | 0 | -9.61 | -10.15 | -9.89 | -0.26 | -1.26 | 1.79 |

**Supplementary material 5.** Predicted binding affinities for all scores avaiable in the program Molegro Virtual Docker (all structures in CDKKi dataset).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PDB access code | Ligand Identification | Ligand Chain | Ligand Number | Ki (nM) | Test Set | Plants Score | MolDock Score | Rerank Score | Interaction Score | Protein Score | Water Score | Internal Score | Electrostatic Score | Electrostatic Long Score | Hbond Score | Ligand Efficiency 1 Score | Ligand Efficiency 3 Score |
| 1E1V | CMG | A | 401 | 12000 | 1 | -61.333 | -95.5564 | -82.4453 | -111.355 | -108.602 | -2.75356 | 15.7988 | 0 | 0 | -2.96992 | -5.30869 | -4.58029 |
| 1E1X | NW1 | A | 401 | 1300 | 0 | -59.853 | -86.2035 | -83.0162 | -107.607 | -104.698 | -2.90852 | 21.4034 | 0 | 0 | -8.92387 | -4.78908 | -4.61201 |
| 1H1S | 4SP | A | 1298 | 6 | 0 | -80.161 | -132.691 | -112.177 | -154.792 | -153.206 | -1.58654 | 22.1008 | 0 | 0 | -7.41587 | -4.73898 | -4.00632 |
| 1JSV | U55 | A | 400 | 2000 | 1 | -57.082 | -85.2962 | -62.0957 | -102.198 | -98.6262 | -3.57181 | 16.9018 | 0 | 0 | -7.46088 | -4.73868 | -3.44976 |
| 1OGU | ST8 | A | 1298 | 2400 | 0 | -87.555 | -130.554 | -122.584 | -163.088 | -147.695 | -15.3936 | 32.534 | 0 | 0 | -6.06403 | -4.83534 | -4.54015 |
| 1PXM | CK5 | A | 500 | 60 | 1 | -80.045 | -120.724 | -103.075 | -123.776 | -122.684 | -1.09256 | 3.05285 | 0 | 0 | -5 | -5.74874 | -4.90834 |
| 1PXN | CK6 | A | 500 | 195 | 0 | 81.365 | -122.547 | -104.91 | -129.11 | -128.698 | -0.412079 | 6.56344 | 0 | 0 | -5 | -5.57032 | -4.76866 |
| 1PXO | CK7 | A | 500 | 2 | 1 | -55.396 | -74.8726 | -64.1598 | -77.5981 | -77.5964 | -0.00172361 | 2.72553 | 0 | 0 | -5.95367 | -5.34804 | -4.58284 |
| 1PXP | CK8 | A | 500 | 220 | 0 | -81.949 | -120.696 | -104.535 | -124.298 | -122.799 | -1.49954 | 3.6023 | 0 | 1.34627 | -2.5 | -5.24764 | -4.545 |
| 1PYE | PM1 | A | 700 | 386 | 1 | -66.875 | -139.302 | -109.345 | -133.393 | -133.393 | 0 | -5.90912 | 0 | 0 | -1.96927 | -4.97507 | -3.90516 |
| 1V1K | 3FP | A | 299 | 35000 | 1 | -60.846 | -92.3692 | -79.1929 | -122.993 | -120.232 | -2.761 | 30.6237 | 4.12617 | -0.672297 | -2.94114 | -2.79907 | -2.39978 |
| 2CLX | F18 | A | 1299 | 13300 | 0 | -20.755 | -66.864 | -4.60743 | -75.0279 | -75.0154 | -0.0124821 | 8.16383 | 0 | 0 | -5.96477 | -4.179 | -0.287965 |
| 2EXM | ZIP | A | 400 | 78000 | 0 | -46.863 | -68.6428 | -57.7894 | -79.94 | -78.4829 | -1.45703 | 11.2972 | 0 | 0 | -3.53483 | -4.57619 | -3.85262 |
| 2FVD | LIA | A | 299 | 3 | 0 | -78.266 | -130.833 | -118.594 | -157.697 | -157.311 | -0.386057 | 26.8643 | 0 | 0 | -6.93225 | -4.36109 | -3.95313 |
| 2XMY | CDK | A | 500 | 0.11 | 1 | -74.599 | -112.957 | -92.0239 | -146.313 | -146.605 | 0.292668 | 33.3553 | 0 | 0 | -9.23729 | -3.89508 | -3.17324 |
| 2XNB | Y8L | A | 1299 | 149 | 1 | -90.137 | -120.295 | -109.872 | -143.577 | -140.906 | -2.67146 | 23.2824 | 0 | 0 | -4.77476 | -4.45536 | -4.06933 |
| 3BLR | CPB | A | 940 | 3 | 0 | -82.603 | -108.061 | -102.395 | -139.125 | -138.851 | -0.273781 | 31.0635 | 0 | 0 | -3.0262 | -3.85933 | -3.65696 |
| 3DDQ | RRC | A | 299 | 250 | 0 | -43.317 | -124.865 | -106.837 | -136.674 | -132.931 | -3.74313 | 11.8091 | 0 | 0 | -2.07585 | -4.80249 | -4.1091 |
| 3LFN | A27 | A | 299 | 3160 | 0 | -90.31 | -152.354 | -125.542 | -162.888 | -162.888 | 0 | 10.5338 | 0 | 0 | -5.29663 | -5.44122 | -4.48366 |
| 3LFS | A07 | A | 299 | 2500 | 1 | -63.999 | -98.68 | -84.9027 | -112.264 | -112.264 | 0 | 13.5841 | 0 | 0 | -2.95651 | -4.48545 | -3.85921 |
| 3MY5 | RFZ | A | 300 | 65000 | 0 | -53.683 | -80.3906 | -62.8716 | -79.7688 | -79.7688 | 0 | -0.621702 | 0 | 0 | -2.10327 | -4.01953 | -3.14358 |
| 4ACM | 7YG | A | 1302 | 210 | 0 | -86.067 | -126.033 | -117.251 | -159.902 | -155.563 | -4.33832 | 33.8687 | 0 | 0.684652 | -8.23001 | -4.06558 | -3.78228 |
| 4BCK | T3E | A | 1298 | 4 | 0 | -61.012 | -105.284 | -84.3827 | -108.251 | -107.467 | -0.4 | 2.96644 | 0 | 1.97505 | -5.71496 | -3.89942 | -3.12528 |
| 4BCM | T7Z | A | 1297 | 123 | 0 | -66.569 | -123.99 | -104.656 | -138.326 | -135.541 | -2.35912 | 14.3362 | 0 | 2.10047 | -6.21503 | -3.75726 | -3.17138 |
| 4BCN | T9N | A | 1299 | 12 | 0 | -67.172 | -116.406 | -98.0401 | -119.845 | -118.624 | -0.0154671 | 3.43868 | 0 | 2.02752 | -5.03462 | -4.85027 | -4.085 |
| 4BCO | T6Q | A | 1299 | 131 | 0 | -65.961 | -128.254 | -113.227 | -145.547 | -144.809 | 0.483253 | 17.2932 | 0 | 1.3673 | -2.5 | -3.88647 | -3.43111 |
| 4BCP | T3C | A | 1299 | 568 | 0 | -61.72 | -126.241 | -104.898 | -137.405 | -132.434 | -3.76093 | 11.1642 | 0 | 1.40493 | -4.82127 | -4.20802 | -3.49662 |
| 4BCQ | TJF | A | 1296 | 147 | 0 | -47.897 | -88.7716 | -71.5789 | -89.7822 | -85.8241 | -3.57162 | 1.01055 | 0 | 1.6707 | -1.15751 | -2.8636 | -2.309 |
| 4EOP | 1RO | A | 301 | 890 | 0 | -37.147 | -120.958 | -96.0959 | -149.682 | -145.629 | -4.0533 | 28.7236 | 0 | 0 | -2.5 | -5.03993 | -4.004 |
| 4NJ3 | 2KD | A | 301 | 140 | 0 | -121.354 | -194.26 | -162.816 | -219.574 | -219.55 | -0.0233243 | 25.314 | 1.17197 | -2.08137 | -12.5278 | -5.3961 | -4.52265 |
| 5D1J | 56H | A | 4000 | 38 | 0 | -86.356 | -139.671 | -118.448 | -149.955 | -146.665 | -3.28974 | 10.2835 | 0 | 0 | -5 | -5.58684 | -4.73792 |

**Supplementary material 6.** Predicted binding affinities for all scores avaiable in the program AutoDock Vina (all structures in CDKKi dataset).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PDB access code | Ligand Identification | Ligand Chain | Ligand Number | Ki (nM) | Test Set | Affinity Score | Gauss1 Score | Gauss2 Score | Repulsion Score | Hydrophobic Score | Hydrogen Score |
| 1E1V | CMG | A | 401 | 12000 | 1 | -5.2697 | 70.0055 | 1170.23 | 4.31803 | 12.2018 | 1.75729 |
| 1E1X | NW1 | A | 401 | 1300 | 0 | -5.64929 | 59.4883 | 1140.31 | 4.23207 | 11.5249 | 4.17642 |
| 1H1S | 4SP | A | 1298 | 6 | 0 | -7.8505 | 95.3053 | 1556.1 | 3.46468 | 22.1759 | 2.64518 |
| 1JSV | U55 | A | 400 | 2000 | 1 | -4.35183 | 69.1799 | 1006.59 | 5.33506 | 21.0293 | 2.47553 |
| 1OGU | ST8 | A | 1298 | 2400 | 0 | -7.24096 | 85.2123 | 1552.05 | 2.91938 | 28.3169 | 2.50289 |
| 1PXM | CK5 | A | 500 | 60 | 1 | 5.54069 | 92.5034 | 1366.73 | 23.2374 | 34.1898 | 2.70492 |
| 1PXN | CK6 | A | 500 | 195 | 0 | -1.53617 | 95.219 | 1446.5 | 14.1979 | 32.6565 | 3.02374 |
| 1PXO | CK7 | A | 500 | 2 | 1 | 5.22501 | 57.2072 | 936.753 | 18.8885 | 17.477 | 4.31861 |
| 1PXP | CK8 | A | 500 | 220 | 0 | 7.55651 | 96.9676 | 1436.28 | 26.4813 | 37.1779 | 1.92178 |
| 1PYE | PM1 | A | 700 | 386 | 1 | -7.43539 | 94.0482 | 1571.41 | 5.44149 | 54.9212 | 1.75304 |
| 1V1K | 3FP | A | 299 | 35000 | 1 | -6.13059 | 73.0603 | 1546.31 | 2.71741 | 36.5969 | 0.55461 |
| 2CLX | F18 | A | 1299 | 13300 | 0 | -0.64452 | 68.951 | 1079.92 | 12.6191 | 25.3091 | 4.26281 |
| 2EXM | ZIP | A | 400 | 78000 | 0 | -4.61106 | 49.7809 | 938.219 | 2.48548 | 8.71274 | 1.01282 |
| 2FVD | LIA | A | 299 | 3 | 0 | -8.88914 | 101.443 | 1596.72 | 3.49453 | 30.1233 | 3.05264 |
| 2XMY | CDK | A | 500 | 0.11 | 1 | -6.74932 | 101.858 | 1685.39 | 5.5163 | 34.5942 | 2.06369 |
| 2XNB | Y8L | A | 1299 | 149 | 1 | -8.79908 | 86.0699 | 1597.63 | 2.77486 | 34.5887 | 1.58581 |
| 3BLR | CPB | A | 940 | 3 | 0 | -8.85037 | 90.1702 | 1579.69 | 3.1095 | 52.6922 | 0.56869 |
| 3DDQ | RRC | A | 299 | 250 | 0 | -7.45701 | 85.9442 | 1495.59 | 2.13271 | 52.2106 | 0.5931 |
| 3LFN | A27 | A | 299 | 3160 | 0 | -9.73722 | 115.255 | 1734.59 | 5.62158 | 94.2824 | 2.13176 |
| 3LFS | A07 | A | 299 | 2500 | 1 | -8.24858 | 74.7202 | 1308.18 | 2.27036 | 48.1043 | 0.87249 |
| 3MY5 | RFZ | A | 300 | 65000 | 0 | -6.34138 | 44.3764 | 1129.1 | 1.4597 | 25.8307 | 0 |
| 4ACM | 7YG | A | 1302 | 210 | 0 | -8.54771 | 105.631 | 1875.08 | 3.73683 | 28.7444 | 2.56706 |
| 4BCK | T3E | A | 1298 | 4 | 0 | -5.92957 | 64.5825 | 1336.25 | 3.11357 | 21.1744 | 1.7757 |
| 4BCM | T7Z | A | 1297 | 123 | 0 | -6.44482 | 105.222 | 1666.83 | 6.5941 | 32.5795 | 2.28443 |
| 4BCN | T9N | A | 1299 | 12 | 0 | -6.34714 | 87.6819 | 1419.92 | 4.8454 | 32.0198 | 1.82698 |
| 4BCO | T6Q | A | 1299 | 131 | 0 | -7.40426 | 109.749 | 1820.58 | 5.64226 | 34.0817 | 0.81695 |
| 4BCP | T3C | A | 1299 | 568 | 0 | -6.75877 | 91.7017 | 1705.51 | 5.53666 | 27.988 | 1.62297 |
| 4BCQ | TJF | A | 1296 | 147 | 0 | -5.93318 | 44.2745 | 1270.57 | 0.85869 | 16.8601 | 0.33072 |
| 4EOP | 1RO | A | 301 | 890 | 0 | -8.74318 | 103.329 | 1578.64 | 4.58129 | 61.5311 | 2 |
| 4NJ3 | 2KD | A | 301 | 140 | 0 | -11.5867 | 148.894 | 2188.99 | 7.56375 | 121.688 | 5.41844 |
| 5D1J | 56H | A | 4000 | 38 | 0 | -7.41179 | 83.662 | 1410.65 | 1.7796 | 21.4573 | 0.85682 |

**Supplementary material 7.** Predicted binding affinities for Taba score with 3 variables (d ≤ 4.5 Å) (all structures in CDKKi dataset).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| PDB access code | Ligand Identification | Ligand Chain | Ligand Number | Ki (nM) | Test Set | Experimental Log(Ki) | Taba (Predicted Log(Ki)) |
| 1E1V | CMG | A | 401 | 12000 | 1 | -4.921 | -7.063 |
| 1E1X | NW1 | A | 401 | 1300 | 0 | -5.886 | -7.914 |
| 1H1S | 4SP | A | 1298 | 6 | 0 | -8.222 | -9.302 |
| 1JSV | U55 | A | 400 | 2000 | 1 | -5.699 | -7.403 |
| 1OGU | ST8 | A | 1298 | 2400 | 0 | -5.62 | -8.321 |
| 1PXM | CK5 | A | 500 | 60 | 1 | -7.222 | -7.219 |
| 1PXN | CK6 | A | 500 | 195 | 0 | -6.71 | -7.396 |
| 1PXO | CK7 | A | 500 | 2 | 1 | -8.699 | -7.602 |
| 1PXP | CK8 | A | 500 | 220 | 0 | -6.658 | -7.087 |
| 1PYE | PM1 | A | 700 | 386 | 1 | -6.413 | -7.662 |
| 1V1K | 3FP | A | 299 | 35000 | 1 | -4.456 | -7.192 |
| 2CLX | F18 | A | 1299 | 13300 | 0 | -4.876 | -8.353 |
| 2EXM | ZIP | A | 400 | 78000 | 0 | -4.108 | -6.953 |
| 2FVD | LIA | A | 299 | 3 | 0 | -8.523 | -8.269 |
| 2XMY | CDK | A | 500 | 0.11 | 1 | -9.959 | -8.832 |
| 2XNB | Y8L | A | 1299 | 149 | 1 | -6.827 | -7.566 |
| 3BLR | CPB | A | 940 | 3 | 0 | -8.523 | -8.319 |
| 3DDQ | RRC | A | 299 | 250 | 0 | -6.602 | -7.21 |
| 3LFN | A27 | A | 299 | 3160 | 0 | -5.5 | -7.968 |
| 3LFS | A07 | A | 299 | 2500 | 1 | -5.602 | -6.791 |
| 3MY5 | RFZ | A | 300 | 65000 | 0 | -4.187 | -7.456 |
| 4ACM | 7YG | A | 1302 | 210 | 0 | -6.678 | -8.021 |
| 4BCK | T3E | A | 1298 | 4 | 0 | -8.398 | -12.45 |
| 4BCM | T7Z | A | 1297 | 123 | 0 | -9.238 | -6.91 |
| 4BCN | T9N | A | 1299 | 12 | 0 | -7.921 | -8.196 |
| 4BCO | T6Q | A | 1299 | 131 | 0 | -6.883 | -8.393 |
| 4BCP | T3C | A | 1299 | 568 | 0 | -6.246 | -8.696 |
| 4BCQ | TJF | A | 1296 | 147 | 0 | -6.833 | -7.193 |
| 4EOP | 1RO | A | 301 | 890 | 0 | -6.051 | -7.552 |
| 4NJ3 | 2KD | A | 301 | 140 | 0 | -6.854 | -9.854 |
| 5D1J | 56H | A | 4000 | 38 | 0 | -7.42 | -6.852 |