Table 1: structures of compounds with their IC50 and binding affinity

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Compound | IC50 | Binding Affinity |
| 1 |  | 3.39 | -9.5 |
| 2 |  | 3.37 | -8.2 |
| 3 |  | 3.13 | -10 |
| 4 |  | 2.07 | -9.1 |
| 5 |  | 3.47 | -9.7 |
| 6 |  | 2.08 | -9.9 |
| 7 |  | 3.01 | -9.6 |
| 8 |  | 1.36 | -9.4 |
| 9 |  | 2.52 | -9.7 |
| 10 |  | 1.48 | -9.8 |
| 11 |  | 1.37 | -10 |
| 12 |  | 2.78 | -9.6 |
| 13 |  | 1.58 | -8.3 |
| 14 |  | 1.3 | -8.3 |
| 15 |  | 1.02 | -9.4 |
| 16 |  | 1.27 | -10.1 |
| 17 |  | 1.43 | -9.9 |
| 18 |  | 4.12 | -10.2 |
| 19 |  | 7.98 | -10.5 |
| 20 |  | 2.03 | -8.8 |
| 21 |  | 8.28 | -9.5 |
| 22 |  | 3.82 | -8.4 |
| 23 |  | 2.5 | -10.2 |
| 24 |  | 2.43 | -8.6 |
| 25 |  | 2.42 | -9.7 |
| 26 |  | 4.27 | -10.4 |
| 27 |  | 2.91 | -8.6 |
| 28 |  | 1.05 | -8.7 |
| 29 |  | 2.45 | -8.4 |
| 30 |  | 29.77 | -10.3 |
| 31 |  | 2.46 | -8.5 |
| 32 |  | 4.79 | -8.6 |
| 33 |  | 25.31 | -10.8 |
| 34 |  | 6.6 | -9.9 |
| 35 |  | 2.39 | -10.5 |
| 36 |  | 5.69 | -10 |
| 37 |  | 7.59 | -9.9 |
|  |  |  |  |
| 38 |  | 18.79 | -9.3 |
| 39 |  | 2.58 | -10.3 |
| 40 |  | 14.67 | -10.3 |
| 41 |  | 9.1 | -9.8 |
| 42 |  | 18.48 | -10.6 |
| 43 |  | 12.01 | -10.4 |
| 44 |  | 0.77 | -9.2 |
| 45 |  | 2.34 | -9.9 |
| 46 |  | 4.94 | -10.2 |
| 47 |  | 3.17 | -8.1 |
| 48 |  | 7.47 | -9.7 |
| 49 |  | 2.76 | -10 |
| 50 |  | 1.76 | -10.1 |
| 51 |  | 1.31 | -9.5 |
| 52 |  | 2.31 | -8.1 |
| 53 |  | 3.97 | -10.4 |
| 54 |  | 1.78 | -10.2 |
| 55 |  | 1.23 | -10 |
| 56 |  | 0.8 | -9.1 |
| 57 |  | 0.00133 | -10.4 |
| 58 |  | 3.04 | -10.1 |
| 59 |  | 1.13 | -10 |
| 60 |  | 1.62 | -10 |
| 61 |  | 0.84 | -10 |
| 62 |  | 2.49 | -8.9 |
| 63 |  | 1.82 | -10.2 |
| 64 |  | 0.85 | -10.3 |
| 65 |  | 0.86 | -10 |
| 66 |  | 3.05 | -8.9 |
| 67 |  | 0.014 | -8 |
| 68 |  | 1.85 | -10.2 |
| 69 |  | 2.97 | -7.9 |
| 70 |  | 1.84 | -9.8 |
| 71 |  | 0.75 | -10.3 |
| 72 |  | 2.36 | -9.9 |
| 73 |  | 1.3 | -8.2 |
| 74 |  | 3 | -10.1 |
| 75 |  | 0.773 | -10.1 |
| 76 |  | 3.22 | -10.3 |
| 77 |  | 2.98 | -10.3 |
| 78 |  | 3.76 | -8 |
| 79 |  | 2.35 | -10 |
| 80 |  | 4.23 | -8 |
| 81 |  | 2.9 | -10.2 |
| 82 |  | 5.03 | -9.6 |
| 83 |  | 4.31 | -10.1 |
| 84 |  | 4.61 | -10.3 |
| 85 |  | 3.1 | -9.8 |
| 86 |  | 4.98 | -9.8 |
| 87 |  | 3.58 | -10.4 |
| 88 |  | 2.96 | -10.2 |
| 89 |  | 8.33 | -9.6 |
| 90 |  | 4.9 | -9.9 |
| 91 |  | 2.71 | -10.1 |
| 92 |  | 1.19 | -10.1 |
| 93 |  | 2.68 | -9.9 |
| 94 |  | 1.92 | -9.8 |
| 95 |  | 1.46 | -10.1 |
| 96 |  | 4.04 | -9.8 |
| 97 |  | 2.81 | -9.9 |
| 98 |  | 4.75 | -10.3 |
| 99 |  | 3.57 | -10 |
| 100 |  | 4.49 | -9.2 |
| 101 |  | 2.23 | -9.8 |
| 102 |  | 4.4 | -10.4 |
| 103 |  | 2.54 | -10.2 |
| 104 |  | 1.74 | -7.7 |
| 105 |  | 1.61 | -7.7 |
| 106 |  | 1.64 | -7.9 |
| 107 |  | 3.34 | -10.1 |
| 108 |  | 3.89 | -10.2 |
| 109 |  | 2.95 | -10.4 |
| 110 |  | 2.84 | -9.8 |
| 111 |  | 4.54 | -11.1 |
| 112 |  | 4.33 | -11.1 |
| 113 |  | 3.25 | -10.9 |
| 114 |  | 3.49 | -8.7 |
| 115 |  | 2.82 | -10.9 |
| 116 |  | 1.67 | -8.2 |
| 117 |  | 2.1 | -8.1 |
| 118 |  | 2.69 | -9.9 |
| 119 |  | 2.22 | -10.3 |
| 120 |  | 8.61 | -10.3 |
| 121 |  | 14.16 | -9.2 |
| 122 |  | 6.98 | -10.9 |
| 123 |  | 6.82 | -11 |
| 124 |  | 11.27 | -8.6 |
| 125 |  | 1.32 | -10 |
| 126 |  | 2.62 | -10 |
| 127 |  | 2.77 | -10 |
| 128 |  | 4.77 | -9.3 |
| 129 |  | 2.63 | -10.3 |

Table 2 external validation result

|  |  |  |
| --- | --- | --- |
| Model biasness test | Systematic Error Result | Absent |
|  | R^2Test (100% data) | 0.8524 |
|  | R0^2Test (100% data) | 0.8513 |
|  | R0^'2Test (100% data) | 0.8374 |
| Classical Metrics | Q2F1(100% data) | 0.8431 |
| (for 100% data) | Q2F2(100% data) | 0.8430 |
|  | Scaled Avg.Rm^2(100% data) | 0.7784 |
|  | Scaled DeltaRm^2(100% data) | 0.0990 |
|  | CCC (100% data) | 0.9190 |
|  | R^2Test (95% data) | 0.9012 |
|  | R0^2Test (95% data) | 0.8436 |
| Classical Metric | R0^'2Test (95% data) | 0.8032 |
| (after removing | Q2F1(95% data) | 0.8173 |
| 5% data with | Q2F2(95% data) | 0.8092 |
| high residuals) | ScaledAvgRm2(95% data) | 0.8599 |
|  | ScaledDeltaRm2(95% data) | 0.0722 |
|  | CCC (95% data) | 0.9229 |
| Number of test set compounds | N. Comp. Test | 38 |
| Range and Mean (train and test) | Train range | 1.8031 |
|  | Train(Y) Mean | 4.1849 |
|  | Test range | 2.1959 |
|  | Test Y Mean | 4.1933 |
|  | (0.1\*Training Set Range) | 0.1803 |
| Threshold values utilized | (0.15\*Training Set Range) | 0.2705 |
| to judge the model predictions | (0.2\*Training Set Range) | 0.3606 |
|  | (0.25\*Training Set Range) | 0.4508 |
| RESULT (MAE-based criteria applied on 95% data) | Prediction Quality | GOOD |

Table 4: Y-Randomization

|  |  |  |  |
| --- | --- | --- | --- |
| Model | R | R2 | Q2 |
| Original | 0.969877 | 0.940662 | 0.932862 |
| Random 1 | 0.144834 | 0.020977 | -0.07399 |
| Random 2 | 0.11593 | 0.01344 | -0.07296 |
| Random 3 | 0.177837 | 0.031626 | -0.0521 |
| Random 4 | 0.15073 | 0.02272 | -0.05959 |
| Random 5 | 0.16904 | 0.028574 | -0.06362 |
| Random 6 | 0.286063 | 0.081832 | -0.01255 |
| Random 7 | 0.231428 | 0.053559 | -0.03192 |
| Random 8 | 0.123313 | 0.015206 | -0.08255 |
| Random 9 | 0.252212 | 0.063611 | -0.01652 |
| Random 10 | 0.131251 | 0.017227 | -0.07478 |
| Random  Models Parameters   |  |  | | --- | --- | | Average r: | 0.178264 | | Average r2: | 0.034877 | | Average Q2: | -0.05406 | | cRp2: | 0.924636 | |