

**Supporting Information**

**Chemoinformatic Analysis of Combinatorial Libraries,  
Drugs, Natural Products and Molecular Libraries Small  
Molecule Repository**

Narender Singh, Rajarshi Guha, Marc Giulianotti, Clemencia Pinilla, Richard Houghten, and  
Jose L. Medina-Franco

**Table S1.** Molecular similarity of random samples of 1,000 molecules each using different fingerprints**Natural products**

| Random sample | MACCS |       |        | GpiDAPH3 |       |        | TGD   |       |        |
|---------------|-------|-------|--------|----------|-------|--------|-------|-------|--------|
|               | Mean  | Stdev | Median | Mean     | Stdev | Median | Mean  | Stdev | Median |
| 1             | 0.392 | 0.139 | 0.388  | 0.235    | 0.157 | 0.261  | 0.616 | 0.122 | 0.627  |
| 2             | 0.399 | 0.135 | 0.398  | 0.240    | 0.154 | 0.265  | 0.628 | 0.112 | 0.637  |
| 3             | 0.390 | 0.138 | 0.389  | 0.231    | 0.158 | 0.257  | 0.616 | 0.121 | 0.627  |
| 4             | 0.394 | 0.134 | 0.390  | 0.238    | 0.157 | 0.264  | 0.620 | 0.122 | 0.632  |
| 5             | 0.398 | 0.136 | 0.397  | 0.228    | 0.157 | 0.255  | 0.621 | 0.117 | 0.632  |

**MLSMR**

| Random sample | MACCS |       |        | GpiDAPH3 |       |        | TGD   |       |        |
|---------------|-------|-------|--------|----------|-------|--------|-------|-------|--------|
|               | Mean  | Stdev | Median | Mean     | Stdev | Median | Mean  | Stdev | Median |
| 1             | 0.400 | 0.112 | 0.395  | 0.230    | 0.147 | 0.259  | 0.690 | 0.098 | 0.697  |
| 2             | 0.393 | 0.111 | 0.389  | 0.215    | 0.148 | 0.247  | 0.689 | 0.100 | 0.697  |
| 3             | 0.395 | 0.113 | 0.390  | 0.223    | 0.148 | 0.253  | 0.688 | 0.098 | 0.695  |
| 4             | 0.403 | 0.113 | 0.398  | 0.236    | 0.144 | 0.263  | 0.694 | 0.100 | 0.703  |
| 5             | 0.398 | 0.111 | 0.394  | 0.223    | 0.147 | 0.253  | 0.693 | 0.095 | 0.701  |

**Library I**

| Random sample | MACCS |       |        | GpiDAPH3 |       |        | TGD   |       |        |
|---------------|-------|-------|--------|----------|-------|--------|-------|-------|--------|
|               | Mean  | Stdev | Median | Mean     | Stdev | Median | Mean  | Stdev | Median |
| 1             | 0.797 | 0.073 | 0.793  | 0.703    | 0.078 | 0.702  | 0.938 | 0.035 | 0.940  |
| 2             | 0.792 | 0.073 | 0.789  | 0.697    | 0.079 | 0.696  | 0.934 | 0.035 | 0.937  |
| 3             | 0.796 | 0.073 | 0.793  | 0.699    | 0.077 | 0.697  | 0.937 | 0.034 | 0.940  |
| 4             | 0.796 | 0.074 | 0.793  | 0.703    | 0.080 | 0.703  | 0.936 | 0.035 | 0.940  |
| 5             | 0.793 | 0.074 | 0.789  | 0.695    | 0.078 | 0.694  | 0.934 | 0.035 | 0.935  |

**Library II**

| Random sample | MACCS |       |        | GpiDAPH3 |       |        | TGD   |       |        |
|---------------|-------|-------|--------|----------|-------|--------|-------|-------|--------|
|               | Mean  | Stdev | Median | Mean     | Stdev | Median | Mean  | Stdev | Median |
| 1             | 0.778 | 0.078 | 0.774  | 0.746    | 0.068 | 0.746  | 0.926 | 0.061 | 0.945  |
| 2             | 0.779 | 0.079 | 0.776  | 0.747    | 0.069 | 0.747  | 0.930 | 0.061 | 0.946  |
| 3             | 0.774 | 0.079 | 0.770  | 0.741    | 0.070 | 0.741  | 0.928 | 0.062 | 0.946  |
| 4             | 0.777 | 0.078 | 0.774  | 0.747    | 0.068 | 0.747  | 0.929 | 0.061 | 0.946  |
| 5             | 0.779 | 0.079 | 0.776  | 0.747    | 0.068 | 0.747  | 0.931 | 0.062 | 0.946  |

**Library III**

| Random sample | MACCS |       |        | GpiDAPH3 |       |        | TGD   |       |        |
|---------------|-------|-------|--------|----------|-------|--------|-------|-------|--------|
|               | Mean  | Stdev | Median | Mean     | Stdev | Median | Mean  | Stdev | Median |
| 1             | 0.781 | 0.081 | 0.774  | 0.564    | 0.114 | 0.565  | 0.894 | 0.060 | 0.900  |
| 2             | 0.779 | 0.082 | 0.771  | 0.564    | 0.113 | 0.564  | 0.896 | 0.061 | 0.902  |
| 3             | 0.783 | 0.082 | 0.776  | 0.571    | 0.113 | 0.571  | 0.896 | 0.060 | 0.902  |
| 4             | 0.779 | 0.083 | 0.771  | 0.563    | 0.114 | 0.560  | 0.895 | 0.061 | 0.902  |
| 5             | 0.781 | 0.085 | 0.776  | 0.572    | 0.116 | 0.570  | 0.899 | 0.059 | 0.904  |

**Library IV**

| Random sample | MACCS |       |        | GpiDAPH3 |       |        | TGD   |       |        |
|---------------|-------|-------|--------|----------|-------|--------|-------|-------|--------|
|               | Mean  | Stdev | Median | Mean     | Stdev | Median | Mean  | Stdev | Median |
| 1             | 0.782 | 0.073 | 0.776  | 0.651    | 0.085 | 0.642  | 0.932 | 0.036 | 0.933  |
| 2             | 0.783 | 0.074 | 0.776  | 0.646    | 0.087 | 0.637  | 0.932 | 0.037 | 0.933  |
| 3             | 0.782 | 0.074 | 0.778  | 0.651    | 0.087 | 0.642  | 0.933 | 0.036 | 0.933  |
| 4             | 0.783 | 0.074 | 0.778  | 0.651    | 0.086 | 0.642  | 0.933 | 0.036 | 0.936  |
| 5             | 0.783 | 0.073 | 0.778  | 0.649    | 0.085 | 0.640  | 0.933 | 0.038 | 0.936  |

**Table S2.** Properties distribution of entire collections and random samples of 1,000 molecules each**Natural products**

| Property | Entire database |        | Random 1 |        | Random 2 |        | Random 3 |        | Random 4 |        | Random 5 |        |
|----------|-----------------|--------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|
|          | Mean            | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median |
| HBA      | 3.28            | 3.00   | 3.23     | 3.00   | 3.32     | 3.00   | 3.26     | 3.00   | 3.19     | 3.00   | 3.27     | 3.00   |
| HBD      | 1.16            | 1.00   | 1.19     | 1.00   | 1.29     | 1.00   | 1.13     | 1.00   | 1.15     | 1.00   | 1.18     | 1.00   |
| RB       | 5.41            | 5.00   | 5.46     | 5.00   | 5.47     | 5.00   | 5.34     | 5.00   | 5.61     | 6.00   | 5.34     | 5.00   |
| SlogP    | 2.75            | 2.89   | 2.73     | 2.93   | 2.75     | 2.89   | 2.75     | 2.93   | 2.88     | 2.88   | 2.79     | 2.93   |
| TPSA     | 77.34           | 75.51  | 78.12    | 75.66  | 78.58    | 76.91  | 76.15    | 74.51  | 75.71    | 74.43  | 75.83    | 70.67  |
| MW       | 374.87          | 385.46 | 376.06   | 389.29 | 378.43   | 388.40 | 371.91   | 380.44 | 378.44   | 389.30 | 373.51   | 387.04 |

**MLSMR**

| Property | Entire database |        | Random 1 |        | Random 2 |        | Random 3 |        | Random 4 |        | Random 5 |        |
|----------|-----------------|--------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|
|          | Mean            | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median |
| HBA      | 3.69            | 4.00   | 3.69     | 4.00   | 3.73     | 4.00   | 3.66     | 4.00   | 3.73     | 4.00   | 3.69     | 4.00   |
| HBD      | 1.21            | 1.00   | 1.18     | 1.00   | 1.27     | 1.00   | 1.18     | 1.00   | 1.27     | 1.00   | 1.27     | 1.00   |
| RB       | 6.06            | 6.00   | 5.97     | 6.00   | 6.00     | 6.00   | 5.93     | 6.00   | 6.09     | 6.00   | 6.11     | 6.00   |
| SlogP    | 3.18            | 3.21   | 3.15     | 3.20   | 3.23     | 3.31   | 3.19     | 3.22   | 3.19     | 3.17   | 3.21     | 3.28   |
| TPSA     | 76.75           | 75.19  | 76.49    | 74.92  | 77.21    | 75.01  | 76.15    | 74.58  | 77.85    | 73.97  | 75.98    | 74.88  |
| MW       | 360.85          | 358.49 | 360.24   | 356.88 | 362.44   | 359.80 | 360.83   | 360.42 | 364.65   | 360.69 | 360.91   | 360.15 |

**Library I**

| Property | Entire database |        | Random 1 |        | Random 2 |        | Random 3 |        | Random 4 |        | Random 5 |        |
|----------|-----------------|--------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|
|          | Mean            | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median |
| HBA      | 3.88            | 4.00   | 3.90     | 4.00   | 3.86     | 4.00   | 3.90     | 4.00   | 3.85     | 4.00   | 3.80     | 4.00   |
| HBD      | 3.69            | 4.00   | 3.70     | 4.00   | 3.69     | 4.00   | 3.72     | 4.00   | 3.69     | 4.00   | 3.64     | 4.00   |
| RB       | 13.14           | 13.00  | 13.13    | 13.00  | 13.08    | 13.00  | 13.19    | 13.00  | 13.09    | 13.00  | 13.18    | 13.00  |
| SlogP    | 5.13            | 5.18   | 5.20     | 5.25   | 5.10     | 5.17   | 5.12     | 5.17   | 5.19     | 5.26   | 5.22     | 5.23   |
| TPSA     | 88.45           | 92.92  | 88.75    | 92.92  | 88.24    | 92.92  | 88.87    | 92.92  | 88.08    | 92.92  | 87.10    | 92.92  |
| MW       | 592.47          | 589.83 | 599.20   | 597.81 | 591.09   | 583.83 | 592.30   | 588.24 | 593.79   | 587.92 | 593.02   | 587.90 |

**Library II**

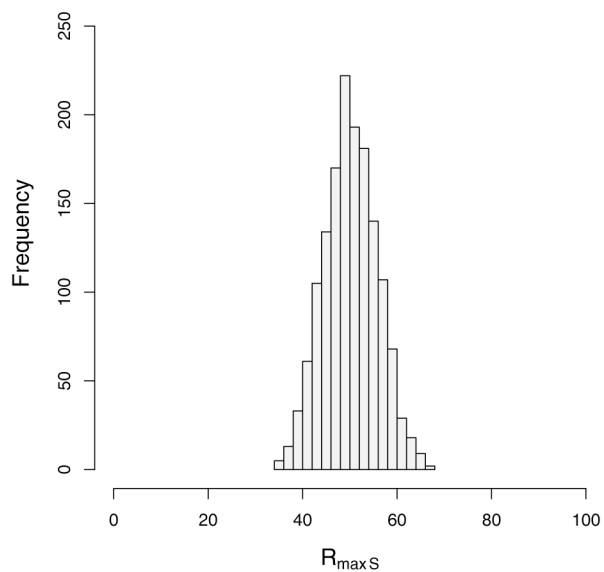
| Property | Entire database |        | Random 1 |        | Random 2 |        | Random 3 |        | Random 4 |        | Random 5 |        |
|----------|-----------------|--------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|
|          | Mean            | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median |
| HBA      | 2.61            | 2.00   | 2.59     | 2.00   | 2.60     | 2.00   | 2.64     | 2.00   | 2.59     | 2.00   | 2.65     | 2.00   |
| HBD      | 3.23            | 3.00   | 3.23     | 3.00   | 3.24     | 3.00   | 3.24     | 3.00   | 3.24     | 3.00   | 3.23     | 3.00   |
| RB       | 13.43           | 13.00  | 13.45    | 13.00  | 13.30    | 13.00  | 13.38    | 13.00  | 13.31    | 13.00  | 13.41    | 13.00  |
| SlogP    | 5.73            | 5.74   | 5.72     | 5.72   | 5.66     | 5.68   | 5.65     | 5.63   | 5.66     | 5.71   | 5.74     | 5.74   |
| TPSA     | 77.64           | 69.45  | 77.37    | 69.45  | 77.63    | 69.45  | 78.04    | 69.45  | 77.50    | 69.45  | 78.00    | 69.45  |
| MW       | 539.12          | 536.74 | 538.25   | 538.78 | 537.54   | 534.73 | 535.66   | 528.83 | 536.88   | 536.83 | 540.91   | 538.83 |

**Library III**

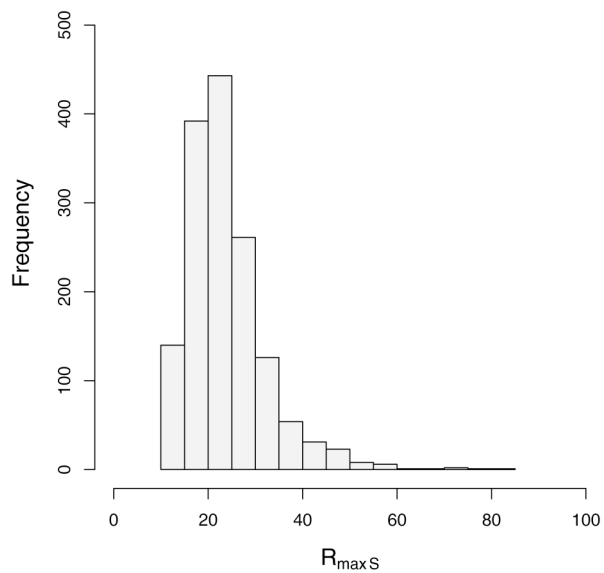
| Property | Entire database |        | Random 1 |        | Random 2 |        | Random 3 |        | Random 4 |        | Random 5 |        |
|----------|-----------------|--------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|
|          | Mean            | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median |
| HBA      | 3.65            | 3.00   | 3.63     | 3.00   | 3.39     | 3.00   | 3.63     | 3.00   | 3.65     | 4.00   | 3.66     | 3.00   |
| HBD      | 1.45            | 1.00   | 1.45     | 1.00   | 1.27     | 1.00   | 1.47     | 1.00   | 1.45     | 1.00   | 1.45     | 1.00   |
| RB       | 9.94            | 10.00  | 9.88     | 10.00  | 9.09     | 9.00   | 9.93     | 10.00  | 9.97     | 10.00  | 9.93     | 10.00  |
| SlogP    | 3.96            | 3.99   | 3.98     | 3.99   | 3.82     | 3.80   | 3.96     | 4.02   | 3.93     | 3.99   | 4.00     | 3.99   |
| TPSA     | 29.19           | 21.75  | 29.03    | 21.75  | 25.14    | 18.51  | 29.18    | 21.75  | 29.10    | 21.75  | 29.29    | 18.51  |
| MW       | 394.30          | 391.69 | 394.53   | 391.67 | 380.77   | 379.59 | 392.28   | 393.62 | 392.58   | 389.63 | 395.90   | 391.69 |

**Library IV**

| Property | Entire database |        | Random 1 |        | Random 2 |        | Random 3 |        | Random 4 |        | Random 5 |        |
|----------|-----------------|--------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|
|          | Mean            | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median | Mean     | Median |
| HBA      | 3.39            | 3.00   | 3.39     | 3.00   | 3.39     | 3.00   | 3.36     | 3.00   | 3.40     | 3.00   | 3.41     | 3.00   |
| HBD      | 2.06            | 2.00   | 2.06     | 2.00   | 2.05     | 2.00   | 2.06     | 2.00   | 2.06     | 2.00   | 2.06     | 2.00   |
| RB       | 8.26            | 8.00   | 8.27     | 8.00   | 8.27     | 8.00   | 8.20     | 8.00   | 8.23     | 8.00   | 8.29     | 8.00   |
| SlogP    | 4.95            | 4.80   | 4.92     | 4.74   | 4.95     | 4.79   | 4.92     | 4.77   | 4.89     | 4.75   | 4.94     | 4.81   |
| TPSA     | 105.48          | 101.73 | 105.14   | 101.73 | 105.30   | 101.73 | 105.46   | 101.73 | 105.10   | 101.73 | 105.68   | 101.73 |
| MW       | 520.90          | 515.31 | 519.64   | 513.66 | 520.41   | 513.65 | 520.33   | 511.60 | 517.80   | 512.19 | 522.89   | 517.65 |



**Figure S1.** Histogram of  $R_{\max(S)}$  values for the DrugBank collection (with respect to itself) using six random descriptors. Each descriptor was generated by sampling from a random uniform distribution, bounded between the maximum and minimum of the corresponding original physicochemical descriptors.



**Figure S2.** Histogram of  $R_{\max(S)}$  values for the DrugBank collection (with respect to itself) using six physicochemical descriptors.