**Table S1.** Comparison of clustering-based selections and diversity-base selections from a random set of 10,000 fragments.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Selected size** | **Selection method** | **Compute time1 (sec)** | **Similarity2** | **Richness2** | **True diversity2** |
| 100 | Clustering-based3 | 42.22 | 0.247 | 1,355 | 605.25 |
| 100 | Diversity-based4 | 17.48 | 0.128 | 2,881 | 1,588.12 |
| 500 | Clustering-based3 | 182.47 | 0.286 | 6,592 | 1,598.00 |
| 500 | Diversity-based4 | 17.39 | 0.166 | 11,273 | 3,342.24 |

1 Computation for compound selections were performed on a Desktop computer (iMac equipped with Intel Core i7 4.2 GHz) using a single processor.

2 There were measured with the same metrics as in the main text.

3 Clustering-based selections were implemented by canvasKMeans of Canvas (Schrödinger, LLC), with commands: “$SCHRODINGER/utilities/canvasKMeans -JOB cl100\_from\_10k -ifp rad64day3\_R10k -k 100 -o cl100” and “$SCHRODINGER/utilities/canvasKMeans -JOB cl500\_from\_10k -ifp rad64day3\_R10k -k 500 -o cl500”.

4 Diversity-based selections were performed as described in the main text, with commands: “$SCHRODINGER/utilities/canvasDBCS -JOB dise100\_from\_10k -ifp rad64day3\_R10k -method dise -metric tanimoto -n 100 -d 0.4 -o dise100.out” and “$SCHRODINGER/utilities/canvasDBCS -JOB dise500\_from\_10k -ifp rad64day3\_R10k -method dise -metric tanimoto -n 500 -d 0.4 -o dise500.out”.