Fragment-based de novo ligand design by multi-objective evolutionary optimization.

Supporting Information

Fabian Dey and Amedeo Caflisch*

Department of Biochemistry
University of Zürich
Winterthurerstrasse 190
CH-8057 Zürich, Switzerland
Phone: (+41 44) 635 55 21
FAX: (+41 44) 635 68 62
email: caflisch@bioc.uzh.ch

* Corresponding author

January 18, 2008

Contents

DAIM-Fingerprint	Average $\pm$ Std.Dev.	Median	Normalization factors
# Atoms	44.99 ± 11.00	45	45
# Carbon	19.17 ± 5.09	19	19
# Nitrogen	2.54 ± 1.45	2	2
# Oxygen	2.88 ± 1.66	3	3
# Halogen	0.70 ± 1.06	0	1
# Sulphur	0.53 ± 0.68	0	1
# Phosphorus	$2.79 * 10^{-3} \pm 0.05$	0	0^a
# Aromatic Bonds	10.69 ± 5.23	12	12
# Double Bonds	3.16 ± 1.69	3	3
# Triple Bonds	0.07 ± 0.28	0	0^a
# Amide Bonds	0.96 ± 0.94	1	1
# Hydrogen Acceptors	3.76 ± 1.67	4	4
# Hydrogen Donors	1.25 ± 1.03	1	1
# Rings	3.13 ± 1.14	3	3
Total Ring Size	16.72 ± 5.64	17	17
Longest Chain	15.90 ± 3.18	16	16
Wiener Index 4 [b]	3.31 ± 1.19	3.21	3.21

Table S1: Average and median of DAIM-fingerprint[2] entries of the ZINC library (version 5).[1] [a] These fingerprint entries were ignored in the 2D fingerprint similarity calculation. [b] Modified Wiener Index $4^{(Ref.3)}$ divided by 1000.

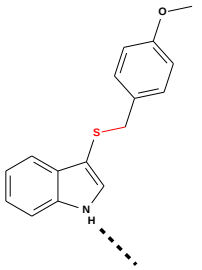
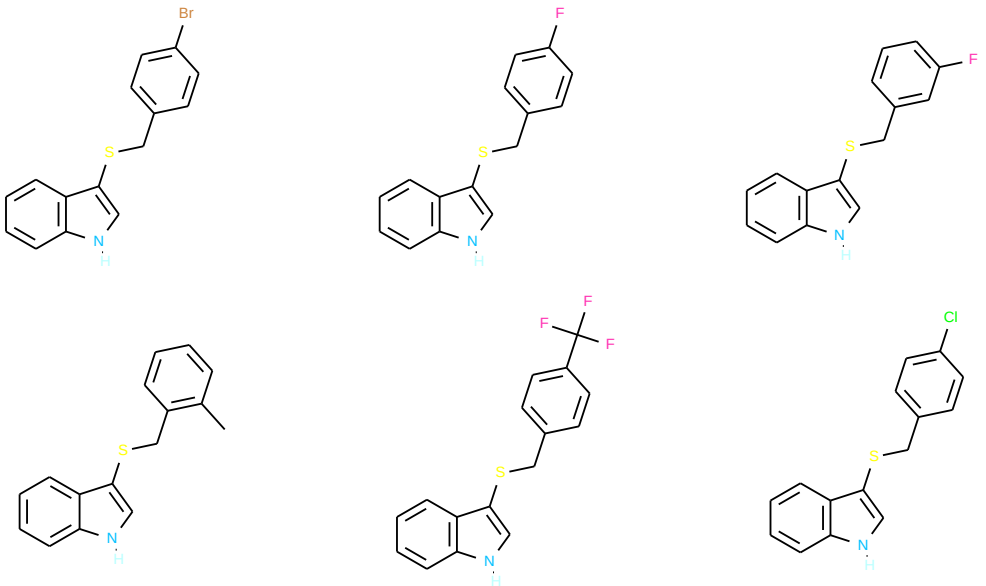
GANDI	
ZINC	

Table S2: Comparison of GANDI molecule from Table 1 with molecules from the ZINC library (version 5)[1] with identical scaffold (a total of 6 molecules were found). The ZINC library was searched with DAIM[2] for molecules containing the query scaffold (ring substituents and apolar hydrogens were removed to yield the query scaffold). In the GANDI molecule, atoms and bonds in red belong to the linker fragment.

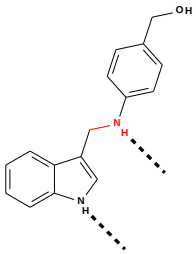
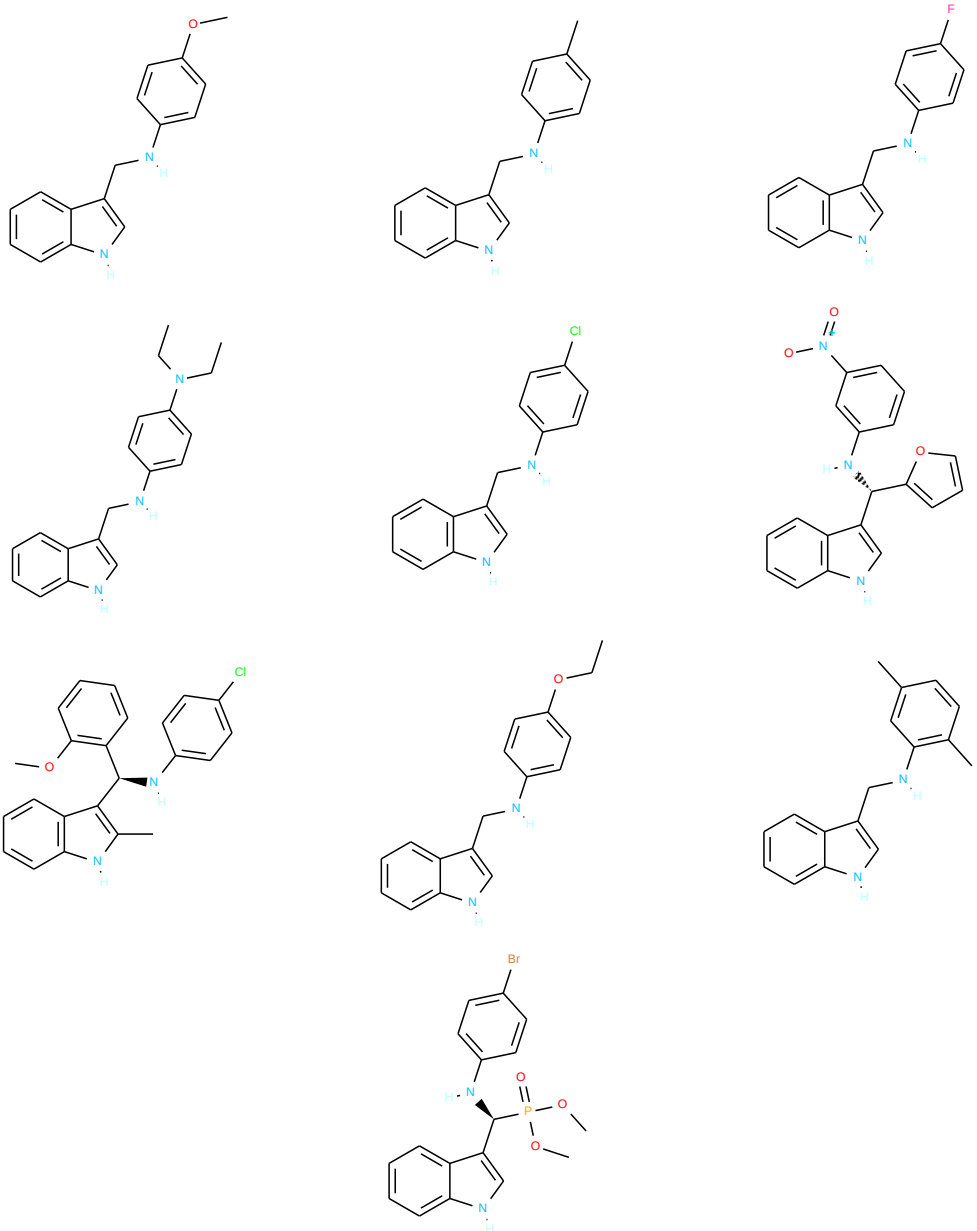
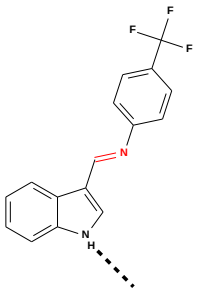
GANDI	
ZINC	

Table S3: Comparison of GANDI molecule from Table 1 with molecules from the ZINC library (version 5)[1] with identical scaffold (a total of 207 molecules were found). See legend of Table S1 for explanations.

GANDI



ZINC

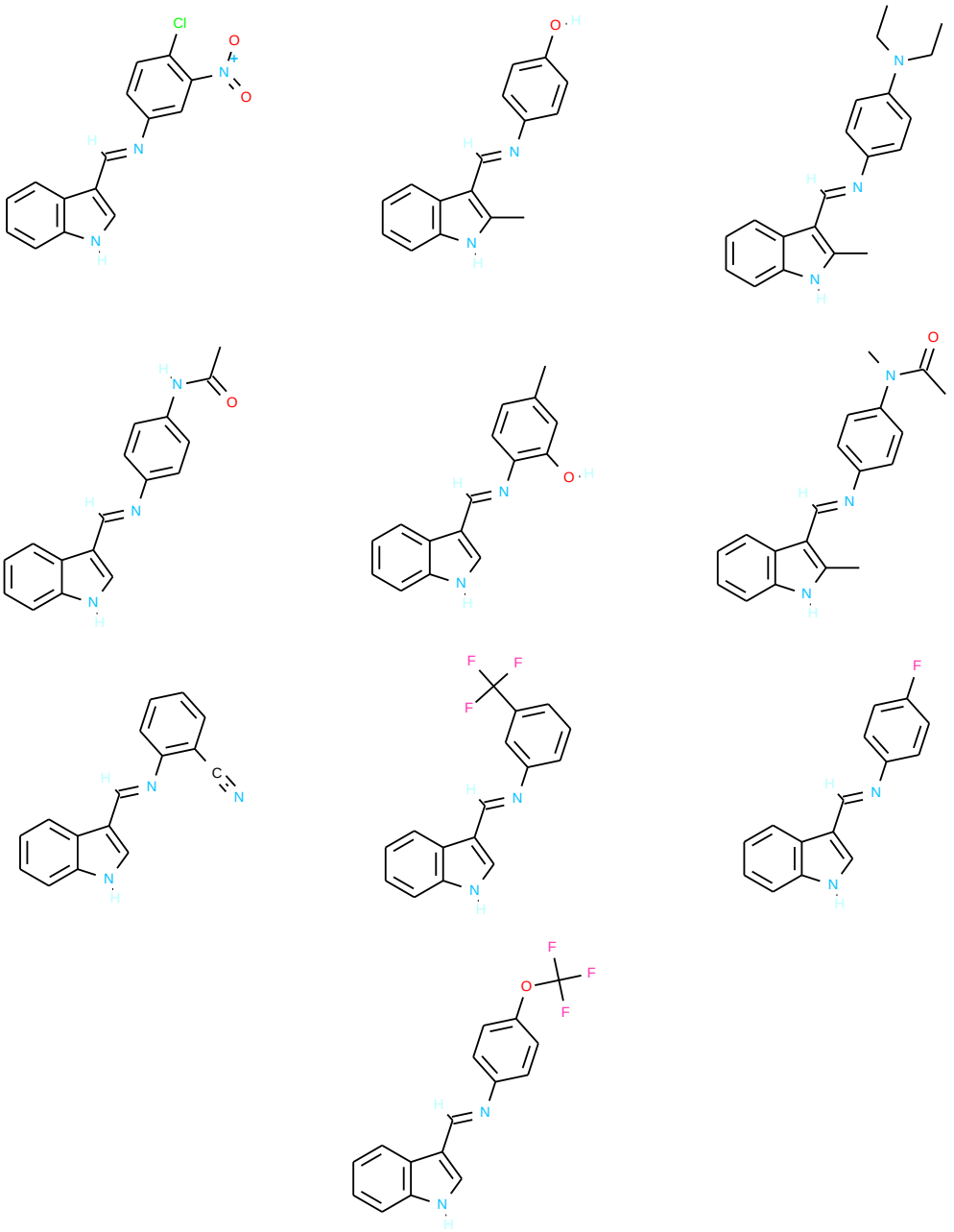
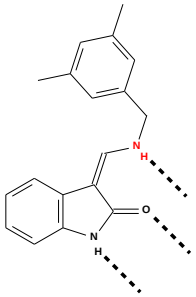


Table S4: Comparison of GANDI molecule from Table 1 with molecules from the ZINC library (version 5)[1] with identical scaffold (a total of 285 molecules were found). See legend of Table S1 for explanations.

GANDI



ZINC

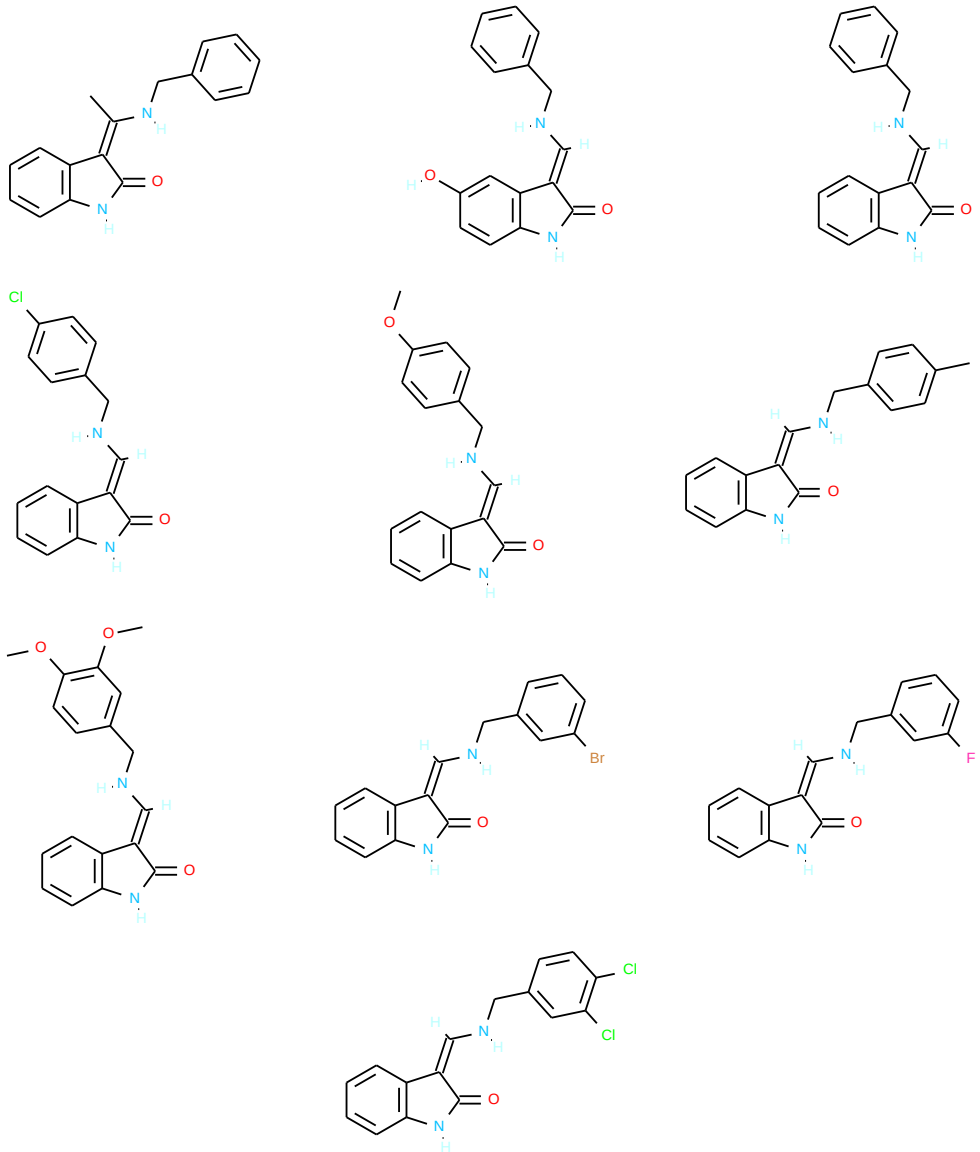
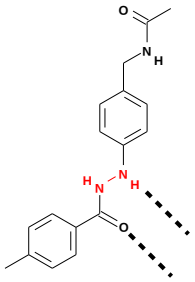


Table S5: Comparison of GANDI molecule from Table 2 with molecules from the ZINC library (version 5)[1] with identical scaffold (a total of 37 molecules were found). See legend of Table S1 for explanations.

GANDI



ZINC

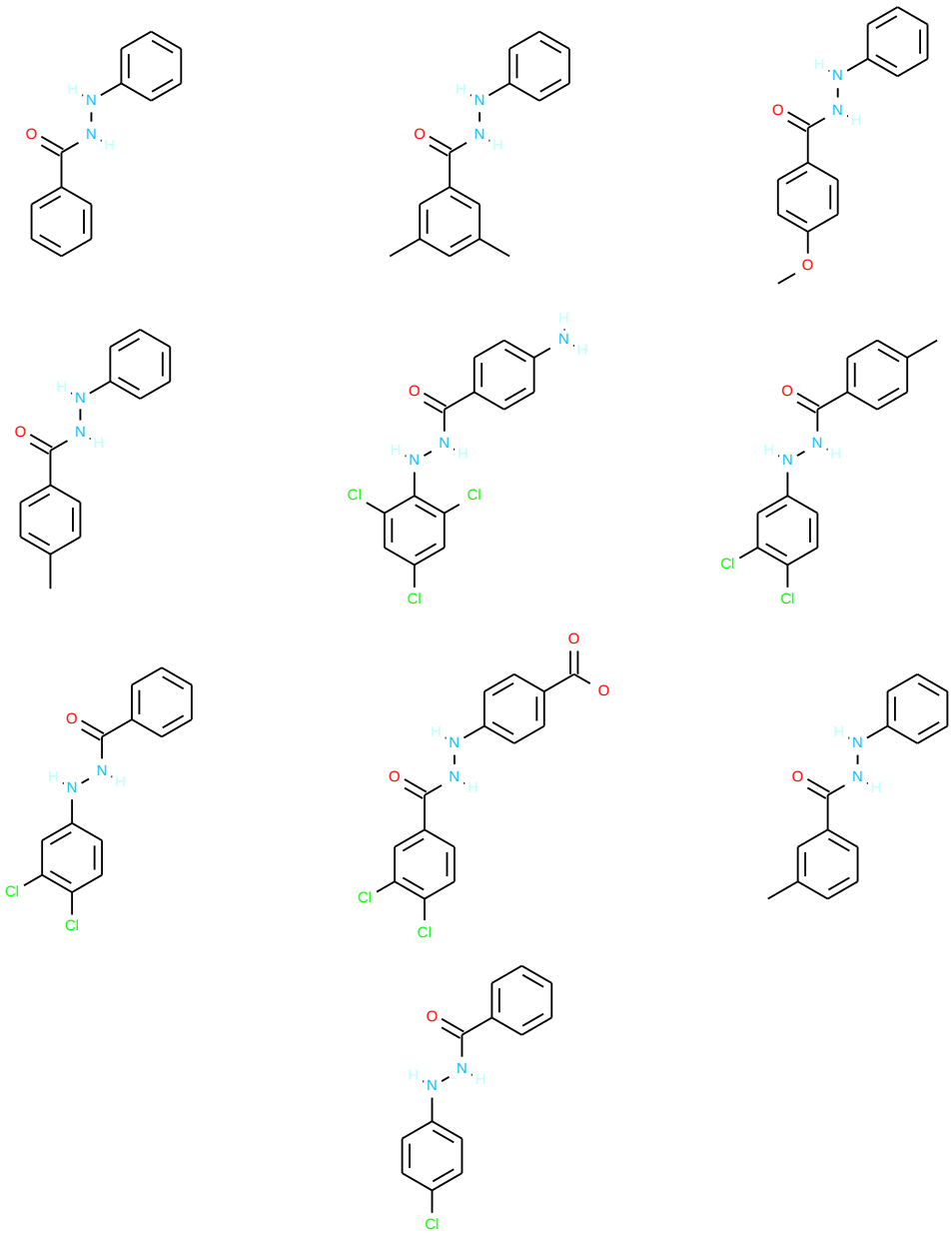


Table S6: Comparison of GANDI molecule from Table 2 with molecules from the ZINC library (version 5)[1] with identical scaffold (a total of 51 molecules were found). See legend of Table S1 for explanations.

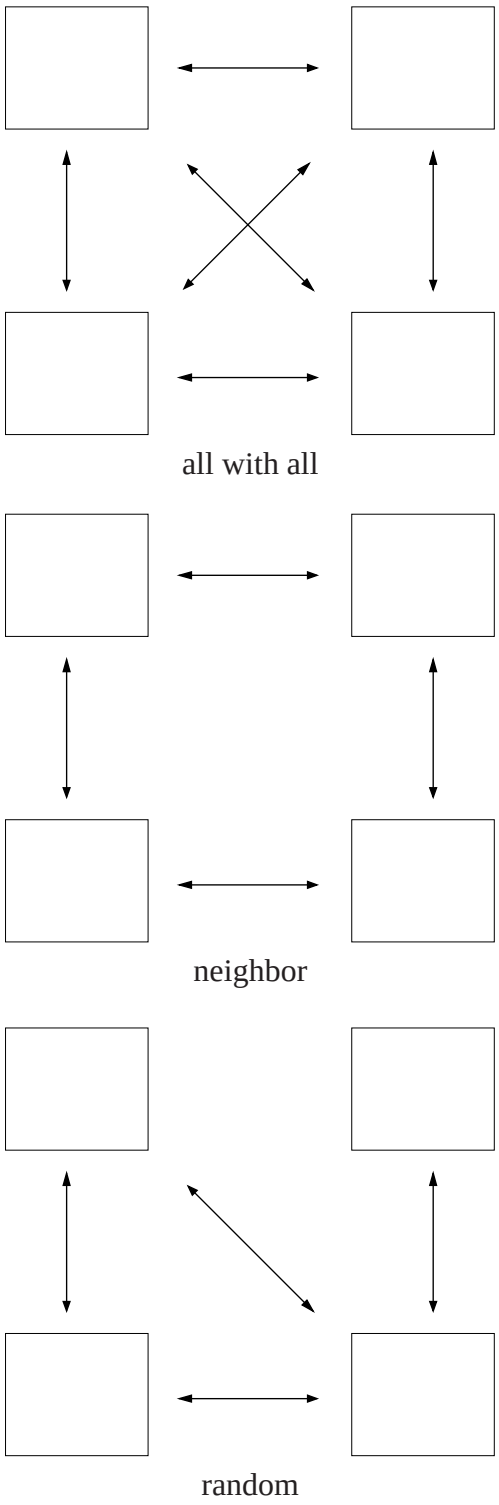


Figure S1: Exchange of molecules between islands of the parallel genetic algorithm: all with all, neighboring or random

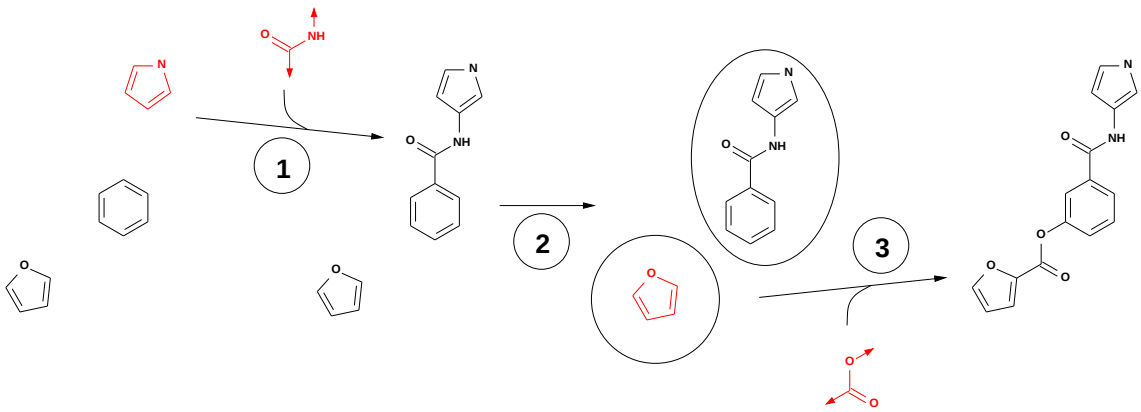


Figure S2: Connecting docked fragments encoded by the genetic algorithm. A suitable linker is selected and the docked fragments are joined in step (1). The fragments are divided into a connected (black) and a not-connected set (red) in step (2). A not-connected docked fragment and a suitable linker are picked randomly to merge the next fragments in step (3). This process is repeated from step (2) until all fragments are connected or a maximal number of trials has been exceeded.

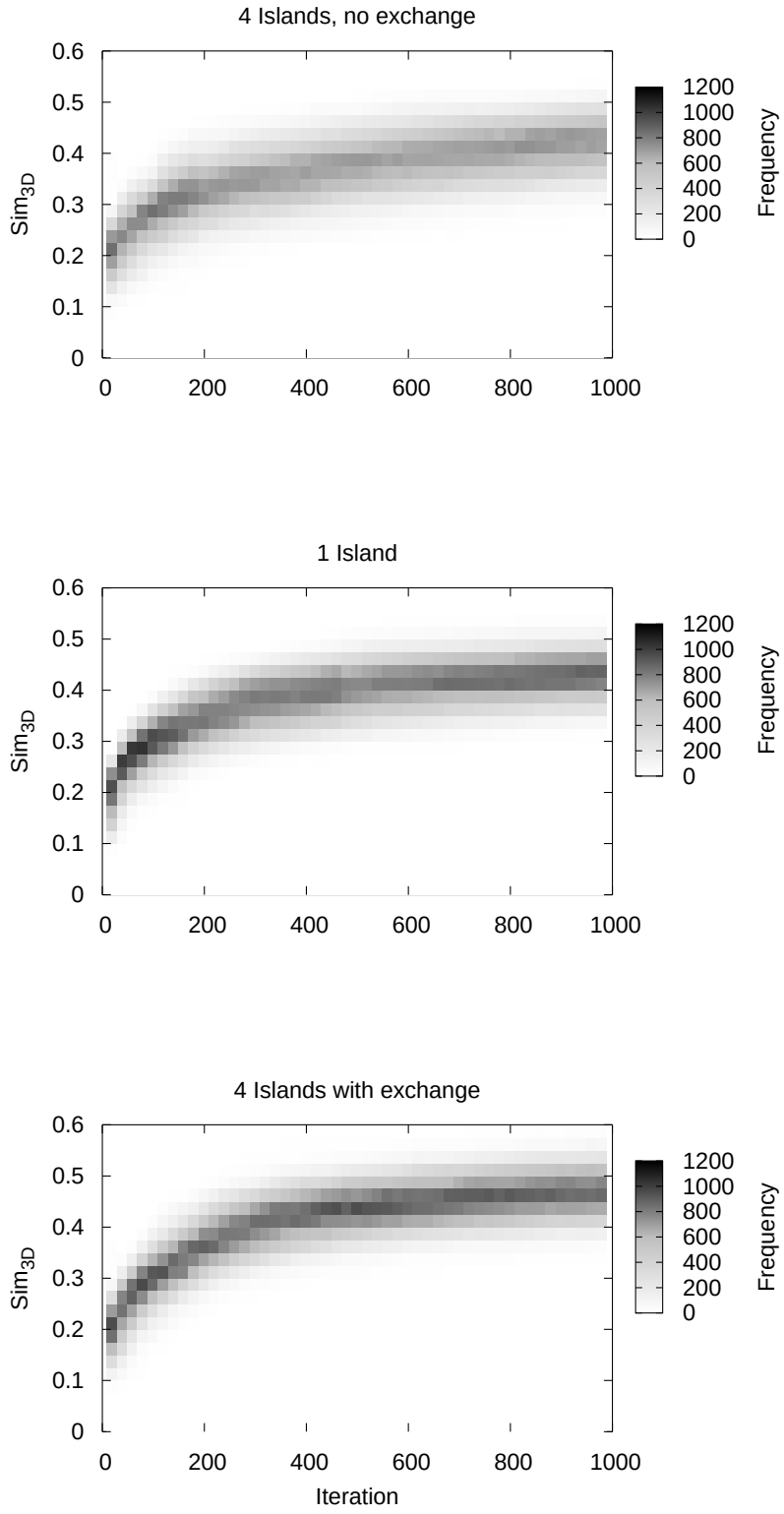


Figure S3: Evolution of the distribution of the pairwise structural similarity (Equation 1), of all pairs of molecules in the same island. Every 20 iteration steps, a vertical histogram is plotted using data from 10 runs.

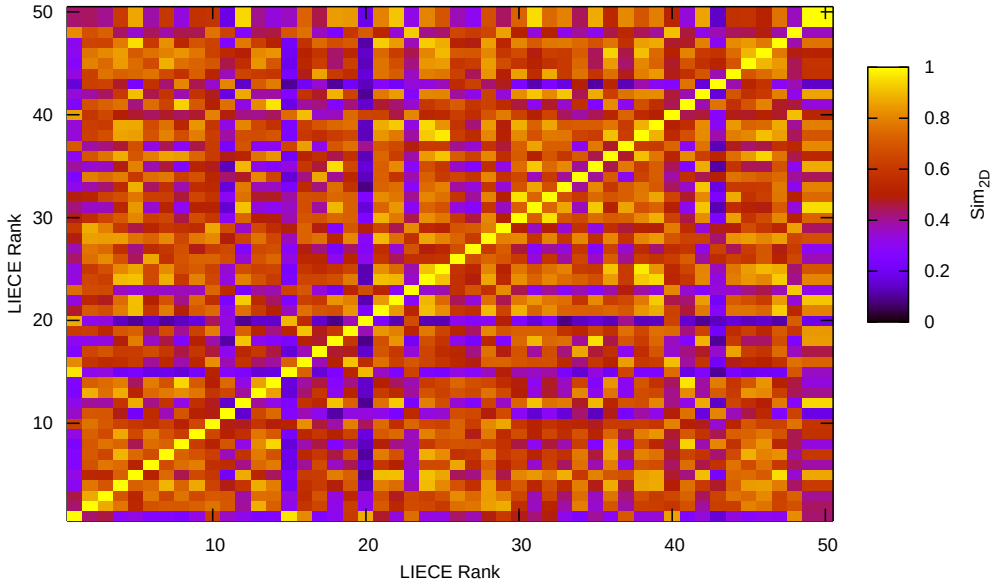


Figure S4: Sim_{2D} similarity matrix of the 50 GANDI molecules with the lowest ΔG_{bind}^{LIECE}. Sim_{2D} was calculated based on normalized DAIM-fingerprints.[2]

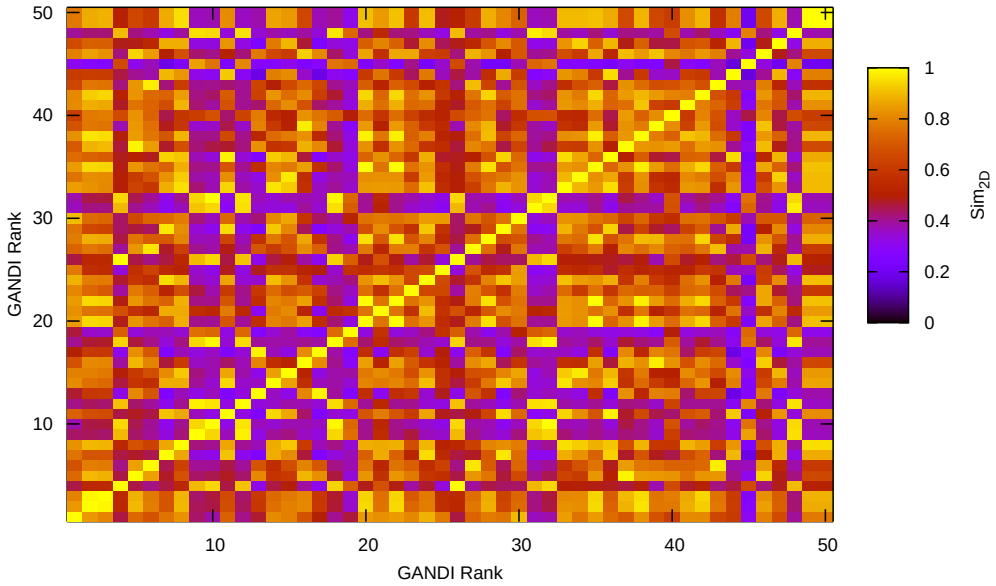


Figure S5: Sim_{2D} similarity matrix of the 50 GANDI molecules with the lowest total GANDI score S_{total}. Sim_{2D} was calculated based on normalized DAIM-fingerprints.[2]

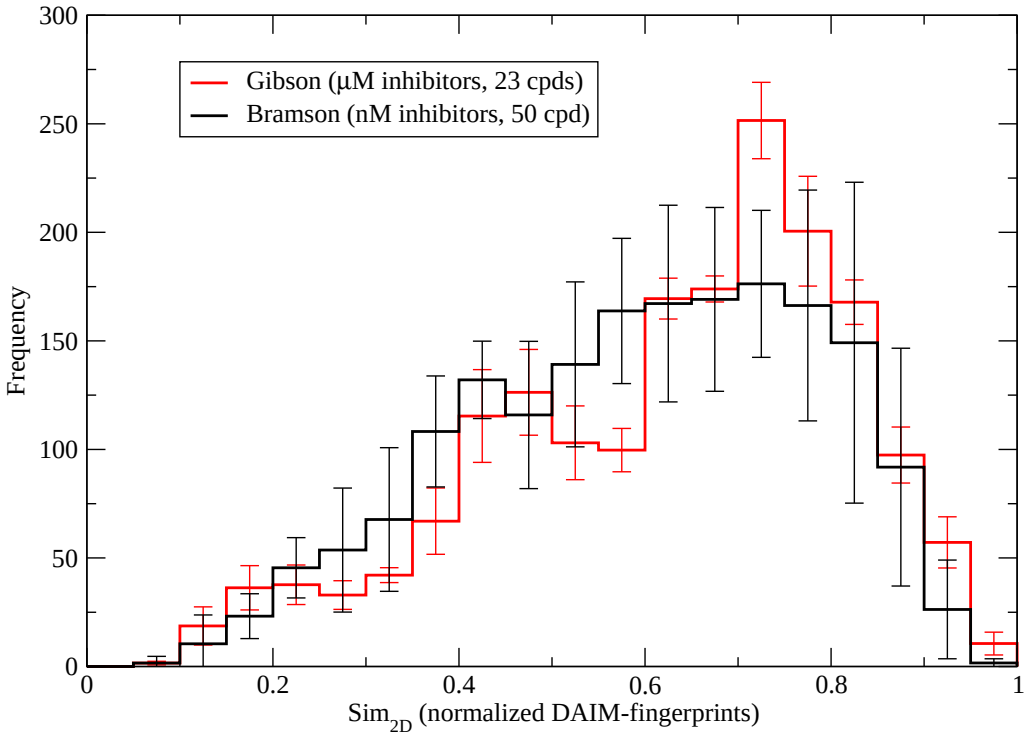


Figure S6: Comparison of known inhibitors with all GANDI molecules. Distribution of Sim_{2D} (Tanimoto similarity,[4] based on normalized DAIM-fingerprints[2]) between a single known inhibitor (from the Bramson[5] and Gibson[6] sets) and the 1'809 GANDI molecules were calculated first. Distributions were then averaged separately for the 23 micromolar (red) and the 50 nanomolar (black) inhibitors. Bars indicate standard deviations.

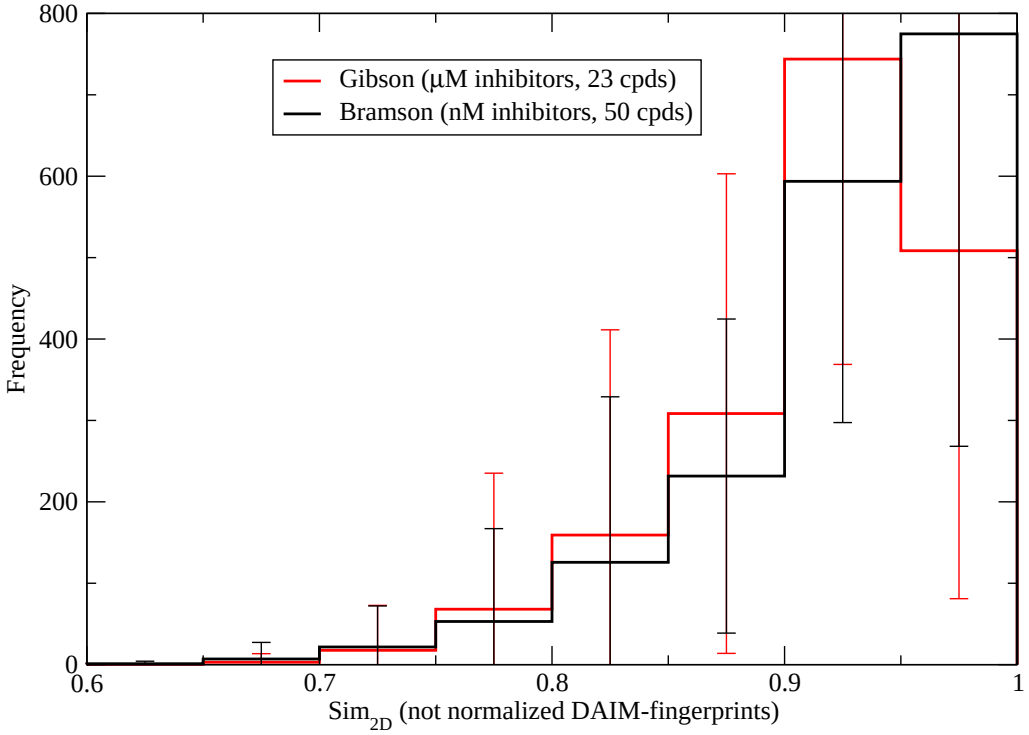


Figure S7: Same as Figure S6 for not-normalized DAIM-fingerprints.[2]

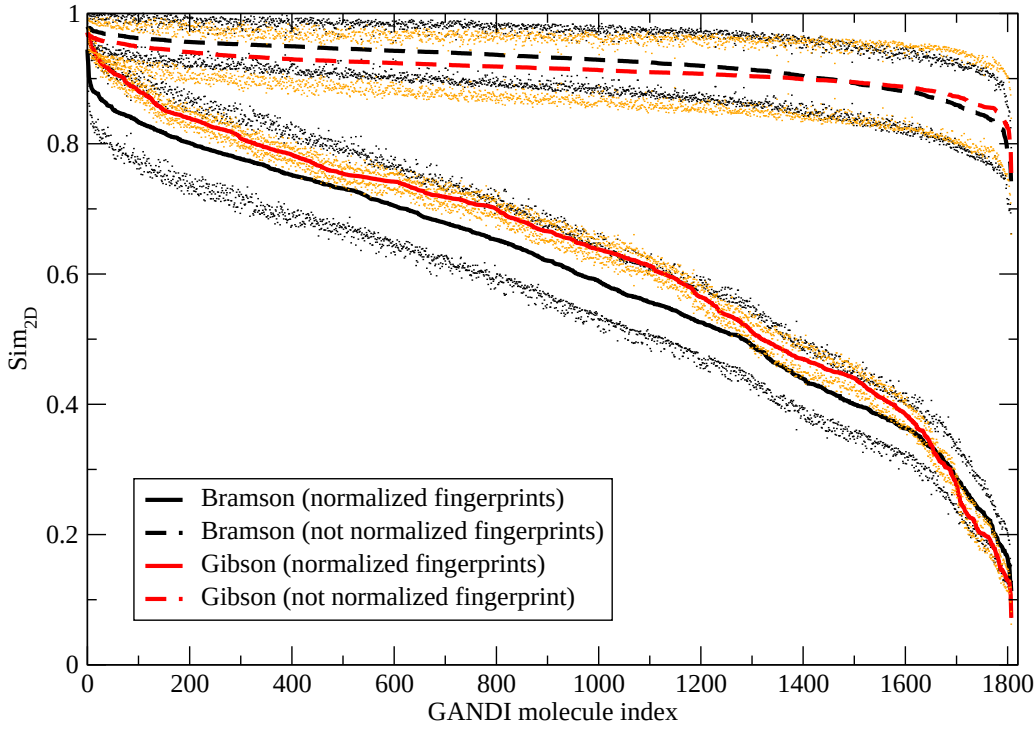


Figure S8: Comparison of single GANDI molecules with known inhibitors. Average similarity Sim_{2D} of every GANDI molecule to the two sets of inhibitors (Bramson[5] and Gibson[6]) based on normalized (solid lines) and not-normalized (dashed lines) DAIM-fingerprints.[2] The standard deviation is given as dots (in orange for the Gibson[6] sets for visibility reasons). The four data sets were sorted individually according to decreasing average similarity.

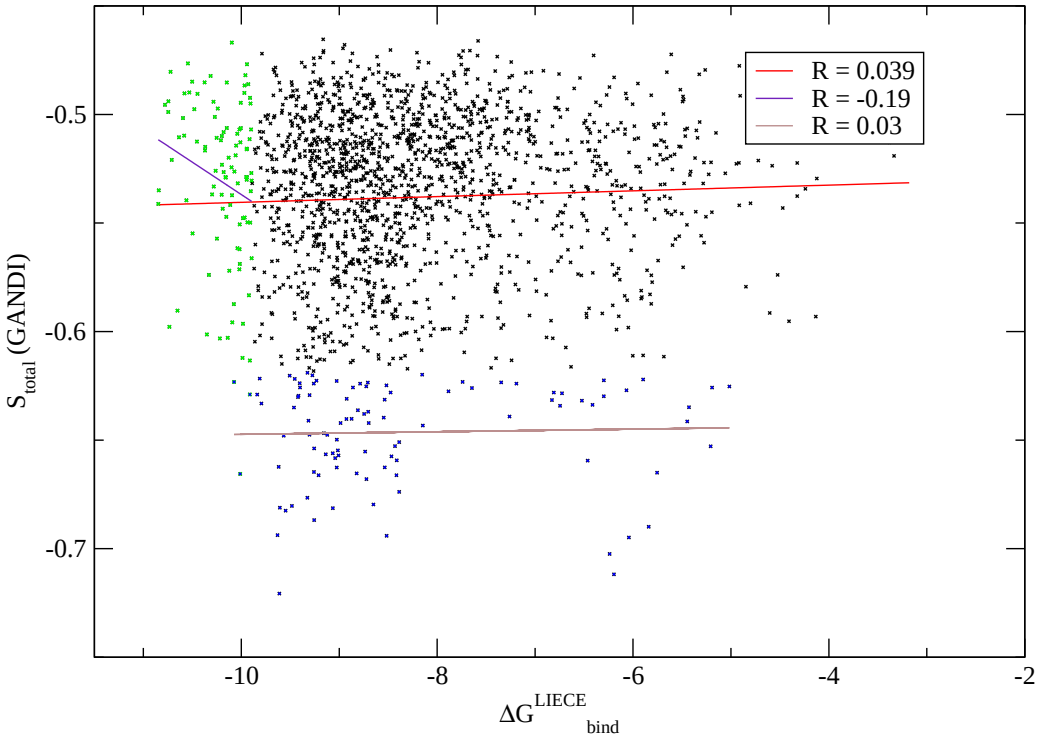


Figure S9: Lack of correlation between ΔG_{bind}^{LIECE} and S_{total} for the 1'809 unique GANDI molecules. Regression analysis was performed on the entire set (red), the 100 GANDI molecules with the lowest ΔG_{bind}^{LIECE} (green dots, magenta line) and the 100 GANDI molecules with the lowest S_{total} (blue dots, brown line).

References

[1] Irwin, J.; Shoichet, B. ZINC – a free database of commercially available compounds for virtual screening., *J. Chem. Inf. Model.* **2005**, *45*, 177–182.

[2] Kolb, P.; Caflisch, A. Automatic and Efficient Decomposition of Two-Dimensional Structures of Small Molecules for Fragment-Based High-Throughput Docking, *J. Med. Chem.* **2006**, *49*, 7384–7392.

[3] Yang, F.; Wang, Z.-D.; Huang, Y.-P. Modification of the Wiener Index 4, *J. Comput. Chem.* **2004**, *25*, 881–887.

[4] Tanimoto, T. *IBM Internal Report*; IBM Technical Report Series 17th November, 1957.

[5] Bramson, H.N.; Corona, J.; Davis, S.T.; Dickerson, S.H.; Edelstein, M.; Frye, S.V.; Gampe, R.T.; Harris, P.A.; Hassell, A.; Holmes, W.D.; Hunter, R.N.; Lackey, K.E.; Lovejoy, B.; Luzzio, M.J.; Montana, V.; Rocque, W.J.; Rusnak, D.; Shewchuk, L.; Veal, J.M.; Walker, D.H.; Kuyper, L.F. Oxindole-Based Inhibitors of Cyclin-Dependent Kinase 2 (CDK2): Design, Synthesis, Enzymatic Activities, and X-ray Crystallographic Analysis, *J. Med. Chem.* **2001**, *44*, 4339–4358.

[6] Gibson, A. E.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Curtin, N. J.; Davies, T. G.; Endicott, J. A.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Whitfield, H. J. Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O6-substituted guanine derivatives, *J. Med. Chem.* **2002**, *45*, 3381–3393.