Bioisostere Effects on the EPSA of Common Permeability-Limiting Groups

Andrew K. Ecker,[1] Dorothy A. Levorse,[2] Daniel A. Victor[2] and Matthew J. Mitcheltree[1,*]

[1] *Department of Discovery Chemistry, Merck & Co., Inc. Boston, Massachusetts, United States*
[2] *Department of Analytical Research & Development, Merck & Co., Inc. Rahway, New Jersey, United States*

* To whom correspondence should be addressed: matthew.mitcheltree@merck.com (M.J.M.)

Table of Contents

No unexpected or unusually high safety hazards were encountered during the execution of this work.
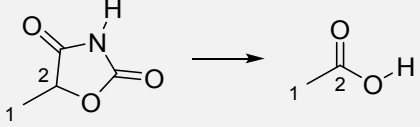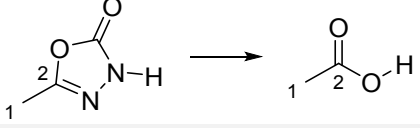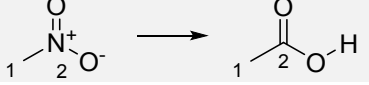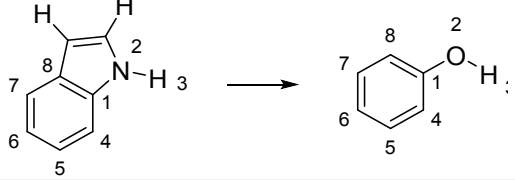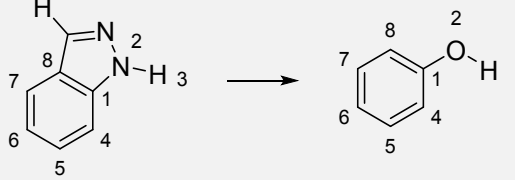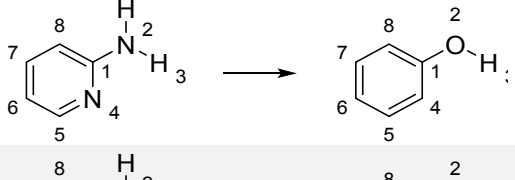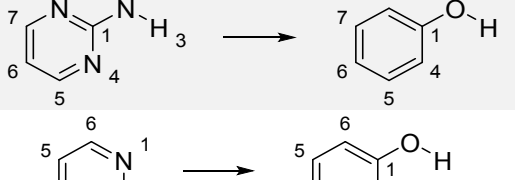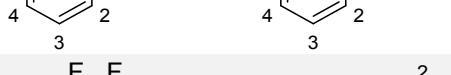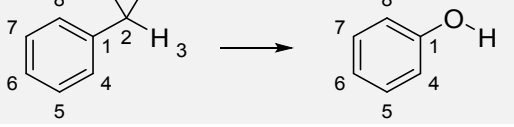
Substructure queries and reaction schemes used in MMP identification

Matched molecular pairs (MMPs) were identified in Pipeline Pilot,[1] using a protocol assembled from widely available components. In it, a substructure query was used to search the corporate collection for compounds containing a given bioisostere substructure. A reaction was performed on each hit, providing a list of all hypothetical MMPs; searching the collection for the enumerated "products" (i.e., structures for which the bioisosteric group is replaced by the parent motif) then provided a list of all actual MMPs in the collection satisfying the relationship parent→bioisostere. **Table S1** lists the substructure queries and reaction schemes used in this workflow, as well as the total number of MMPs identified for each bioisostere type. Also listed are the numbers of MMPs ultimately included in ΔEPSA analysis after filtering for availability and purity, clustering based on N × N 3D similarity (*vide infra*), and manual selection.

Table S1. Reaction schemes used for MMP identification and associated statistics
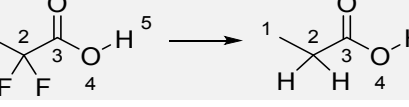
Parent category	Bioisostere type	MMP Transform	Total MMPs (N)	MMPs studied (n)
Amides	Amidine		903	21
Amides	Sulfonamide		19,615	25

Amides	Thioamide	(structure)	1,169	14
Amides	Oxetanyl amine	(structure)	9	5
Amides	1,2,3-Triazole	(structure)	649	25
Amides	1,2,4-Oxadiazole	(structure)	2,704	32
Amides	N-Me Amide	(structure)	38,285	34
Amides	Ester	(structure)	10,051	26
Amides	α-CF$_3$ Amine	(structure)	352	17
Amides	Carbamate	(structure)	6,649	56
Carboxylic acids	α-Difluoro acids	(structure)	92	14
Carboxylic acids	Tetrazole	(structure)	756	25
Carboxylic acids	1,2,4-Oxadiazolone	(structure)	277	9
Carboxylic acids	α-Dimethyl acids	(structure)	510	13

S4

Carboxylic acids	Oxazolidinedione		32	6
Carboxylic acids	1,3,4-Oxadiazolone		347	23
Carboxylic acids	Nitro		6,305	23
Phenols	Indole		954	20
Phenols	Indazole		643	31
Phenols	Aminopyridine		835	23
Phenols	Aminopyrimidine		240	18
Phenols	Pyridine		10,587	20
Phenols	Difluoromethyl benzene		77	19

Clustering by 3D similarity and compound selection

To ensure that for each bioisostere type, a maximally diverse sub-set of MMPs were selected for study, hits from the MMP searches described above were clustered based on three-dimensional shape similarity. From each set of MMPs, the bioisostere-containing structures were first expanded using OMEGA.[2,3] Up to 10 conformations were retained per compound. An N × N 3D-similarity matrix was then computed for all pairs of compounds using the FastROCS Toolkit;[2,4] the highest similarity between conformers of distinct molecules was used for each pair. These 3D similarity scores were used to group the compounds using hierarchical density-based clustering (hdbscan).[5] A typical result is shown in **Figure S1**; in it, colors represent cluster IDs determined by hdbscan, and each point represents a compound (itself representing one half of a unique MMP). Selections were made by choosing compounds (and their MMP partners) from each cluster.
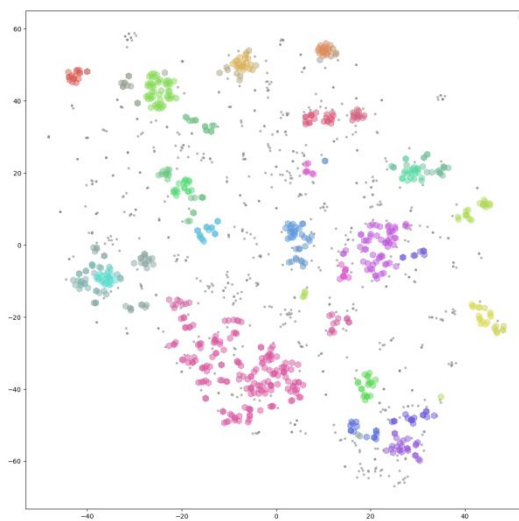
Figure S1. Representative XY plot of 1,2,4-oxadiazoles following HDBSCAN clustering of their N × N 3D similarity (N = 2,704).

Similarity distributions of MMPs included in the analysis

As a measure of the structural diversity represented among the MMPs selected for EPSA analysis, n × n similarity analyses were conducted. For each set of MMPs describing a bioisostere type, n × n Tanimoto similarity matrices were computed using the parent structures of each MMP.[1] The binned frequency distributions of the resulting matrix elements are depicted in **Figures S2– S4**. In each, unity Tanimoto coefficients (similarity = 1, indicating structural identity) correspond to main diagonal matrix elements and thus are inversely proportional in frequency to the square of the sample size, n.

Figure S2. Tanimoto similarity distributions of amide compounds included in the study, grouped by bioisostere type.

Figure S3. Tanimoto similarity distributions of carboxylic acid compounds included in the study, grouped by bioisostere type.
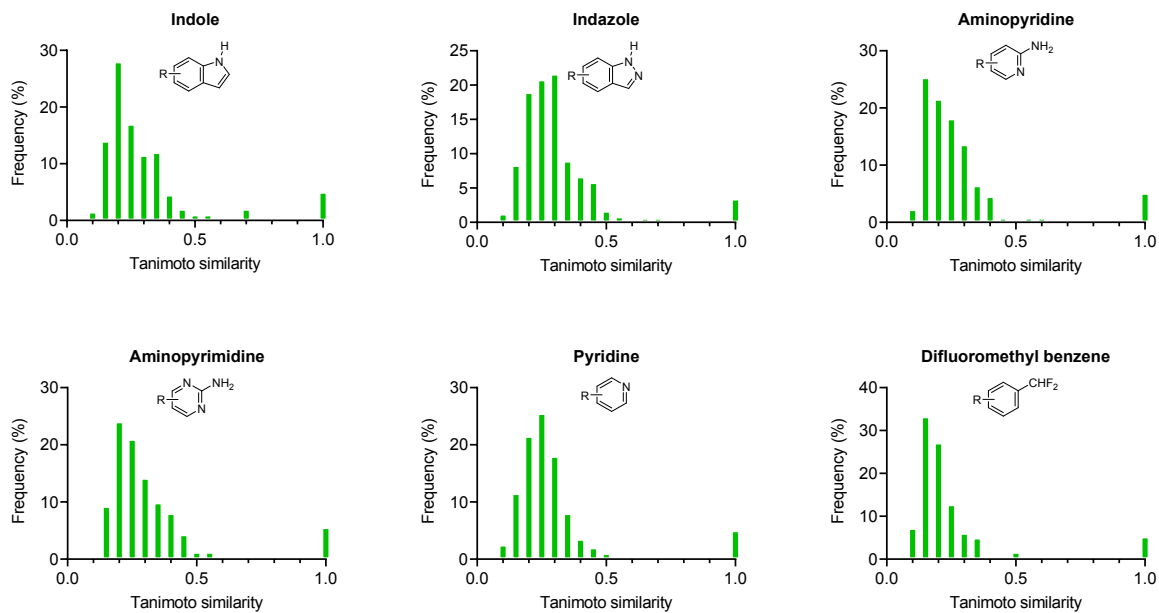
Figure S4. Tanimoto similarity distributions of phenol compounds included in the study, grouped by bioisostere type.

Effect of amidine pK_a on amide→amidine ΔEPSA

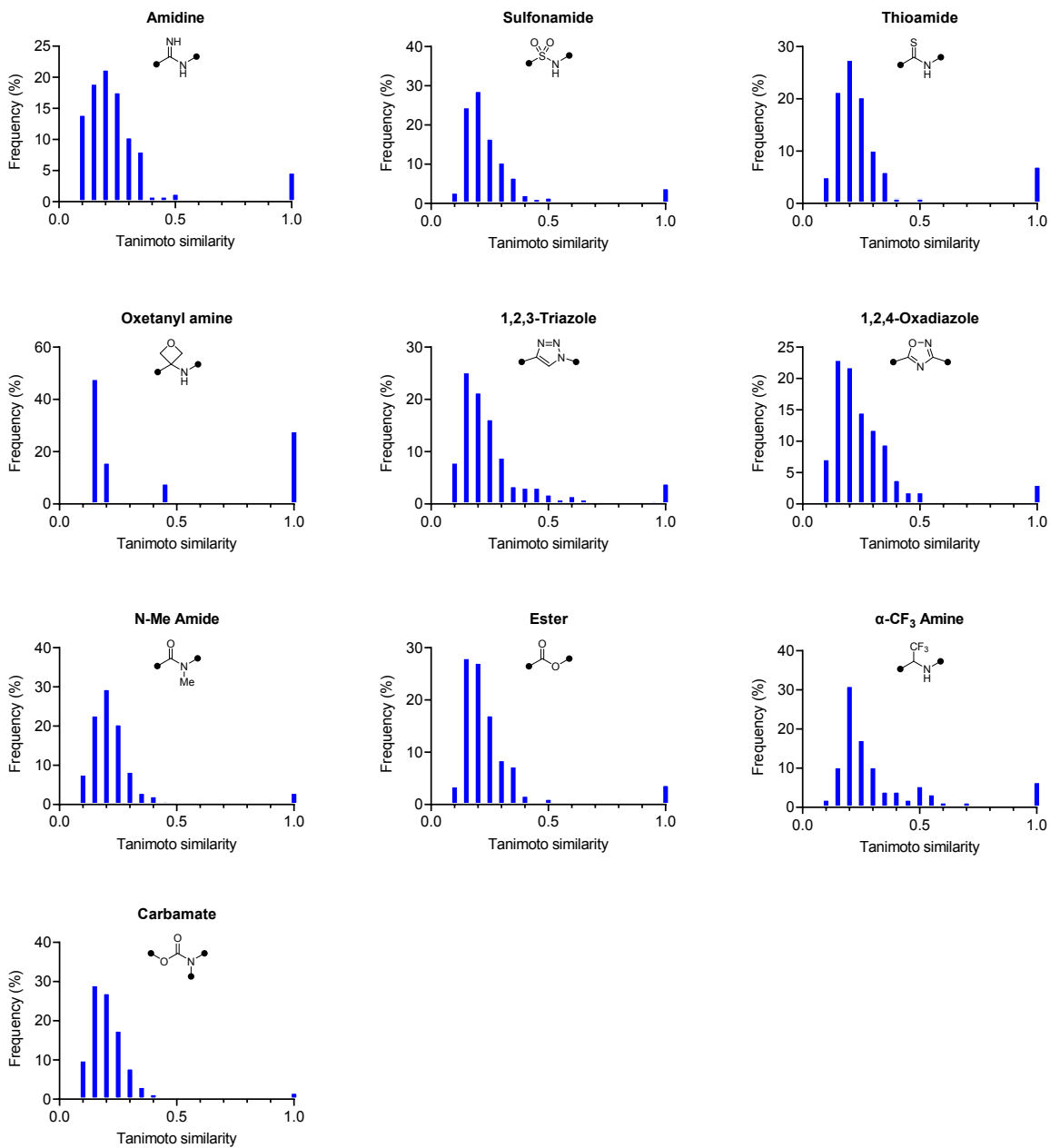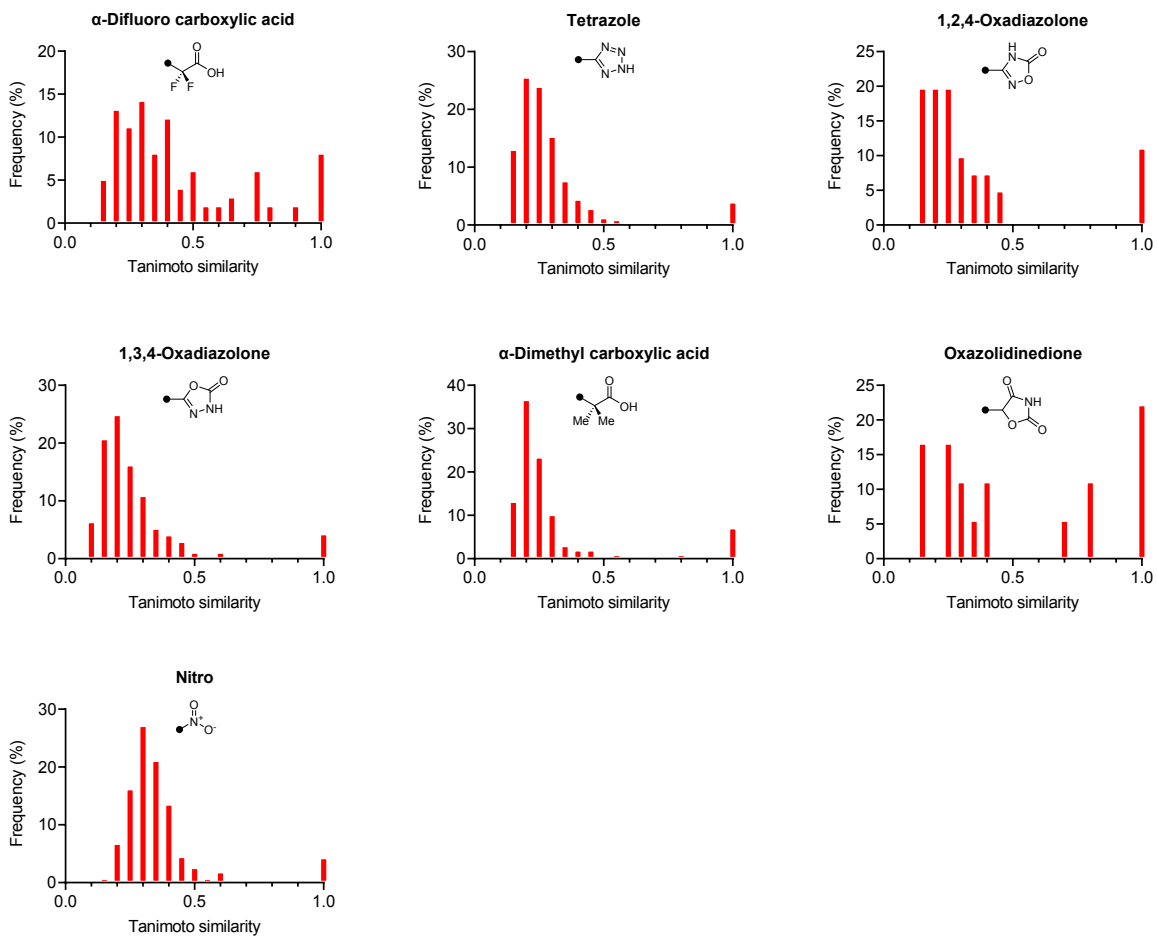To account for the distribution of ΔEPSA values observed upon bioisosteric replacement of amide groups with amidines, the correlation of amidine basicity and ΔEPSA was investigated (**Figure S5**). Within the set of MMPs describing amide→amidine substitution, acid dissociation constants were calculated for the amidine component of each MMP using ACD Classic, ACD GALAS,[6] Epik,[7] and Jaguar.[8] Predictions in Jaguar were performed following thorough conformational searching (the top 5 conformers within a 12.0 kcal/mol energy window were included in the analysis); DFT geometry optimization was performed using the Jaguar implementation of the PBF aqueous solvation model. As expected, MMPs for which the amidine component features greater basicity (i.e., higher predicted pK_a, and thus a greater ionized fraction at neutral pH) showed greater ΔEPSA values. MMPs for which the amidine component was predicted to remain considerably unionized (pK_a < 8) consistently exhibited ΔEPSA < 10.
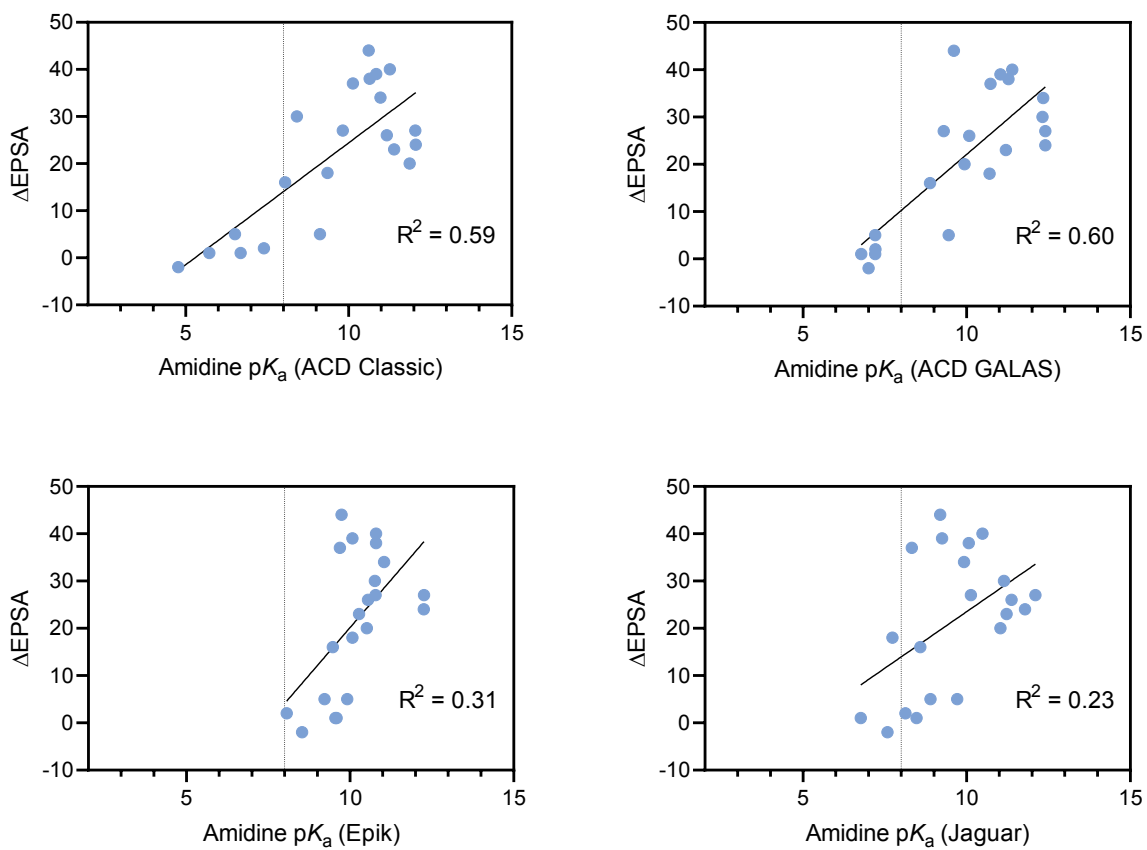
Figure S5. Correlation of amide→amidine ΔEPSA with predicted amidine pK_a.

Statistical deconvolution of structural elements affecting amide→carbamate ΔEPSA

Within the set of MMPs studied describing amide→carbamate bioisosteric substitution, ΔEPSA values appear to follow a bimodal distribution, with a small sub-set exhibiting ΔEPSA > 0. Statistical comparison of MMP sub-sets demonstrated that differences in N-substitution (2° versus 3°) did not significantly affect ΔEPSA (p = 0.41, ns), while the topology of the sub-structure in which the amide or carbamate group is embedded (cyclic versus acyclic) exhibits a strong effect (p = 2.6 ×10^{-9}, ****). The ΔEPSA distributions of these sub-sets are depicted in **Figure S6**. Statistical comparisons were performed using Welch's two-tailed unpaired *t*-test.[9]

Figure S6. ΔEPSA distributions of amide→carbamate MMP sub-sets. Bars and whiskers depict mean ± s.d.

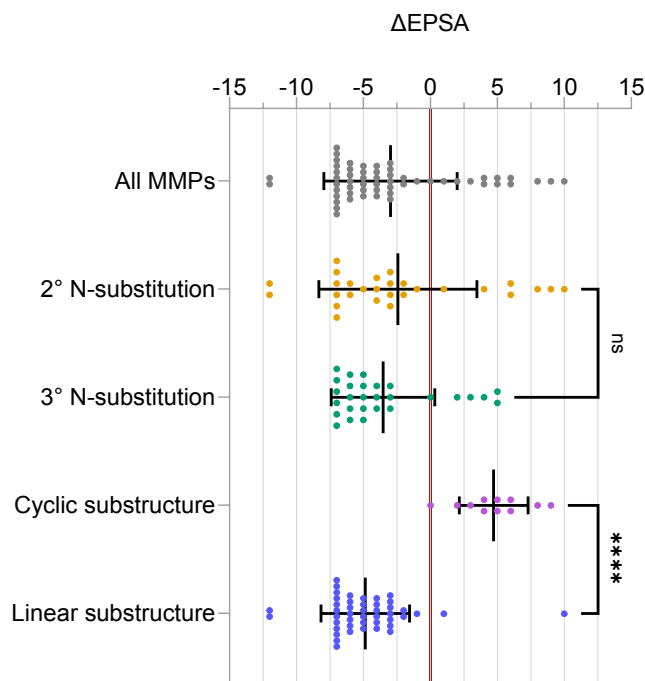**Calculation of Boltzmann-weighted Dipole, HBA basicity, and HBD acidity properties**

Boltzmann-weighted properties for compounds **3**–**6** were calculated using conformational ensembles generated using mixed torsional/low-mode sampling in aqueous solution using a customized OPLS3e force field; all conformers (including mirror-image conformers) within 21.0 kJ/mol of the global minimum were retained.[10] These conformers were then optimized in the gas phase by density functional theory using the B3LYP-D3 functional and 6-311G**$^{++}$ basis set; total Gibbs free energies (used for Boltzmann weighting at T = 298.15 K) and molecular dipoles were computed at this stage. Hydrogen bond donor (N–H) and acceptor (C=O) strengths for each conformer were calculated using Kenny's method[11] as implemented in Jaguar,[8] using the B3LYP functional and LACVP**$^{+}$ basis set. Carbonyl groups typically produced two molecular electrostatic potential minima per conformer; the mean of these two values was used to define HBA basicity for each conformer prior to Boltzmann weighting.

References

1. BIOVIA, Dassault Systèmes. Pipeline Pilot, version 18.1; San Diego: Dassault Systèmes, 2018.

2. OpenEye Scientific Software: Santa Fe, NM, http://www.eyesopen.com

3. Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A., Stahl, M.T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572-584.

4. Grant, J.A.; Gallardo, M.A.; Pickup, B.J. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J. Comp. Chem.* **1996**, *17*, 1653-1666.

5. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software* **2017**, *2*, 11.

6. ACD/Percepta, version 2018.2.1; Advanced Chemistry Development, Inc.: Toronto, ON, Canada, www.acdlabs.com, 2021.

7. Schrödinger Release 2022-1: Epik, Schrödinger, LLC: New York, NY, USA, 2022.

8. Schrödinger Release 2020-1: Jaguar, Schrödinger, LLC: New York, NY, USA, 2020.

9. Prism, version 9.0; GraphPad Software: San Diego, California USA, www.graphpad.com, 2020.

10. Schrödinger Release 2020-1: MacroModel, Schrödinger, LLC: New York, NY, USA, 2020.

11. Kenny, P.W.; Montanari, C.A.; Prokopczyk, I.M.; Ribeiro, J.F.R.; Rartori, G.R. Hydrogen Bond Basicity Prediction for Medicinal Chemistry Design. *J. Med. Chem.* **2016**, *59*, 4278–4288.