Supporting Information for:

Minor Groove Binders and Drugs Targeting Proteins Cover Different Regions in Chemical (Shape) Space

Julian E. Fuchs,[†] Gudrun M. Spitzer,[*,†] Ameera Javed,[†] Adam Biela,[‡] Christoph Kreutz,[†] Bernd Wellenzohn,[§] and Klaus R. Liedl[†]

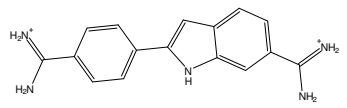
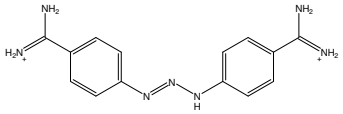
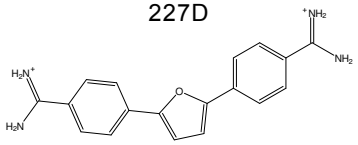
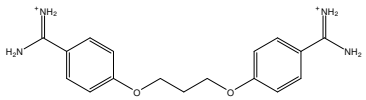
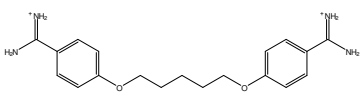
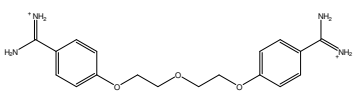
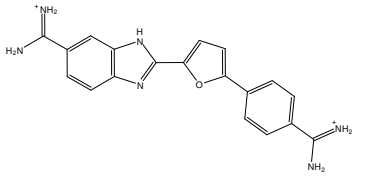
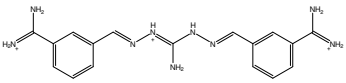
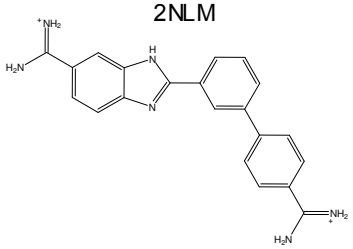
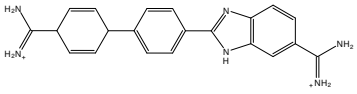
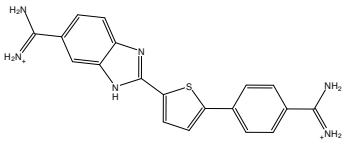
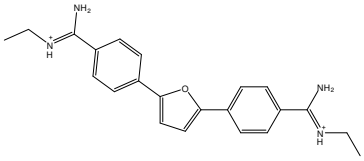
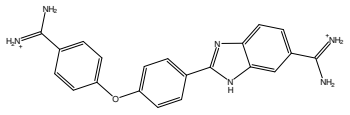
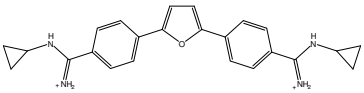
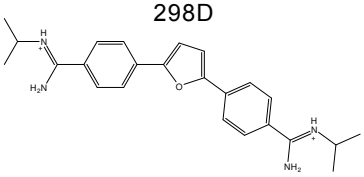
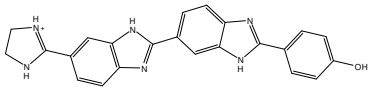
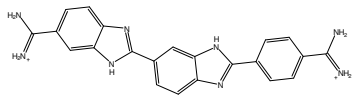
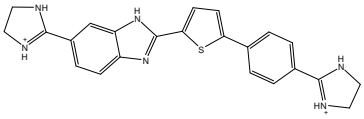
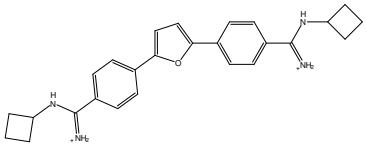
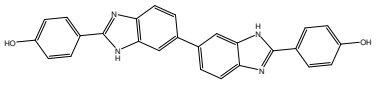
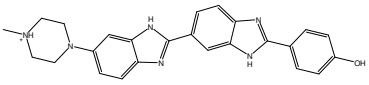
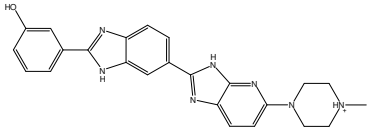
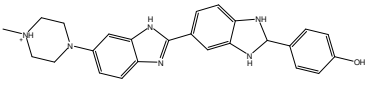
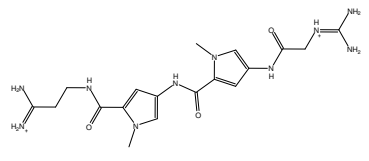
Faculty of Chemistry and Pharmacy, University of Innsbruck, Innrain 52c, A-6020 Innsbruck, Austria, Department of Pharmaceutical Chemistry, Philipps University Marburg, Marbacher Weg 6, D-35032 Marburg, Germany, and Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Strasse 65, D-88397 Biberach/Riss, Germany

E-mail: gudrun.spitzer@uibk.ac.at

DNA Ligands Extracted From PDB Structures

Ligands extracted from crystal structures were used to generate the set of known minor groove binders (MGBs). Following structures have been used: 102D, 109D, 121D, 127D, 129D, 130D, 144D, 166D, 195D, 1D30, 1D43, 1D63, 1D64, 1EEL, 1FMQ, 1FMS, 1FTD, 1JTL, 1LEJ, 1LEX, 1M6F, 1PQQ, 1QV4, 1QV8, 1VZK, 1ZPH, 1ZPI, 227D, 263D, 289D, 298D, 2B0K, 2B3E, 2FJV, 2GYX, 2I2I, 2NLM, 302D, 311D, 328D, 360D, 403D, 442D, 443D, 445D, 449D, 453D

[†]University of Innsbruck
[‡]Philipps University Marburg
[§]Boehringer-Ingelheim

432D 	1D63 	227D 
102D 	1D64 	166D 
2I 2I 	1M6F 	2NLM 
2B0K 	1VZK 	360D 
2FJV 	289D 	298D 
109D 	311D 	2B3E 
1FMQ 	453D 	127D 
302D 	1D43 	121D 

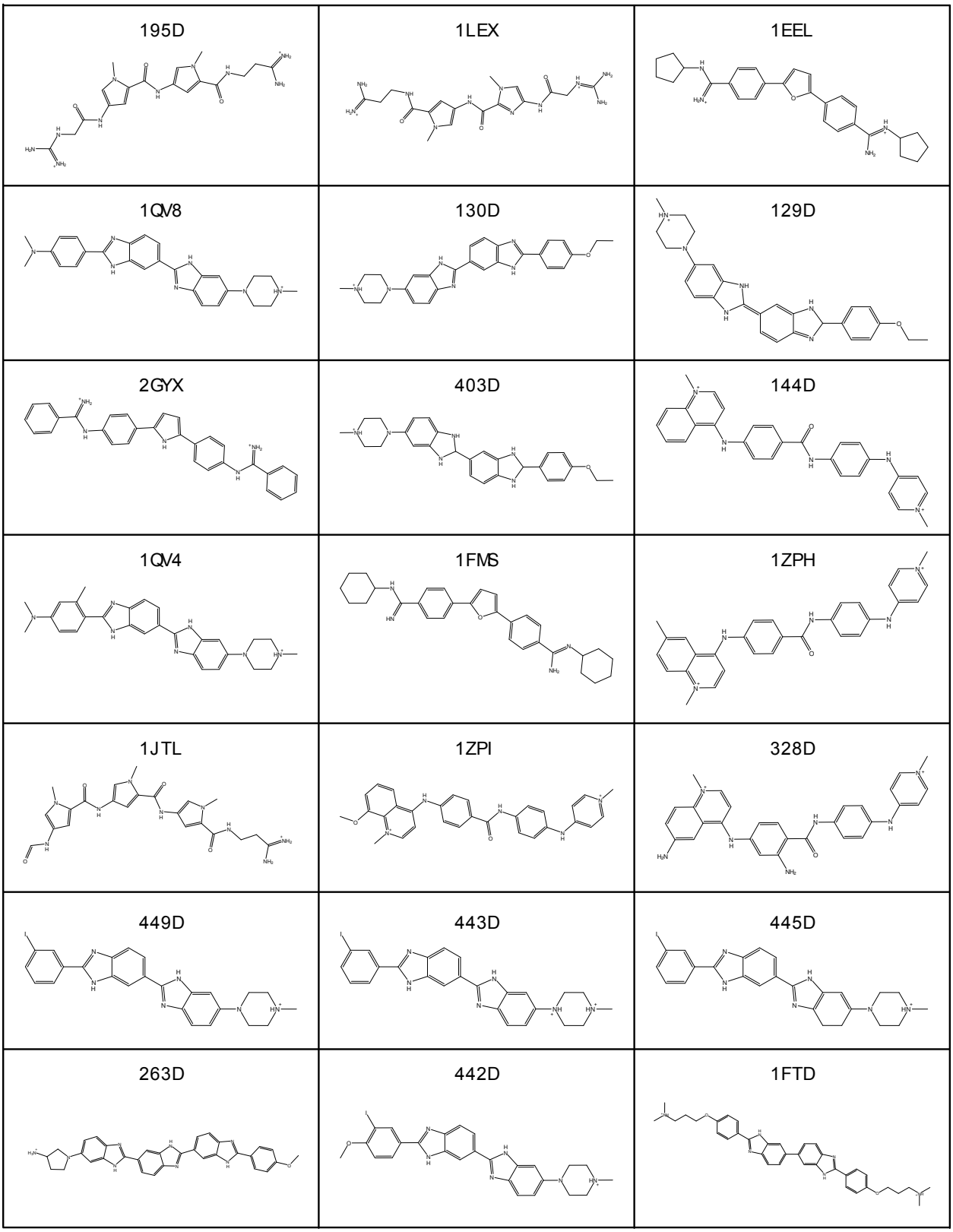


Figure 1: List of ligand structures in the set of known MGBs extracted from crystal structures. PDB codes are indicated for every ligand.

NMR Spectroscopy

The NMR samples were prepared by from lyophilized Dickerson Drew dodecamer DNA. The DNA was dissolved together with 50 μM sodium arsenate buffer, pH 6.5 in H_2O/D_2O = 9/1. Ligands were added from stock solutions (200 μM for ligand **4** and 66 μM for ligand **3**) in 50 μM sodium arsenate buffer, pH 6.5. Data was acquired on a 600 MHz Bruker Avance III NMR instrument at the temperatures indicated. ^{1}H NMR spectra of H_2O-samples were acquired using a double-pulsed field gradient spin-echo (DPFGSE) pulse sequence. The DNA imino resonances were assigned by aid of 2D-^{1}H-^{1}H-NOESY experiments.

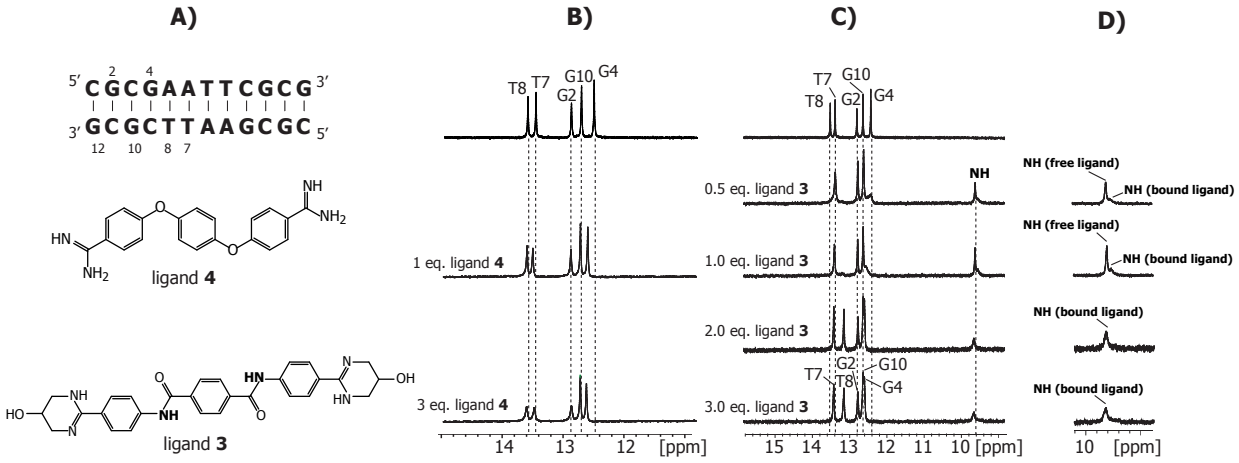


Figure 2: **(A)** Schematic representation of the Dickerson Drew dodecamer (DD DNA), ligand **3** and ligand **4**. **(B)** Monitoring of ligand **4** binding to the DD DNA. Chemical shift perturbation and line broadening effects were most pronounced for the central imino proton resonances (T8, T7 and G4). Thus, ligand **4** binds preferentially at the central tract of the Dickerson Drew dodecamer. **(C)** Monitoring of ligand **3** binding to the DD DNA. Chemical shift perturbation and line broadening effects were again pronounced for the central imino proton resonances (T8, T7 and G4). Furthermore, the amide NH resonance of ligand **3** could be assigned via a NOESY experiment. A weak NOE between T7 NH and the amide proton was also observed. **(D)** Magnification of the ligand NH region. At lower ligand concentrations (0.5 and 1.0 eq. ligand) an equilibrium between free form and the binary complex can be observed. The sharp NH resonance at 9.8 ppm arises from the free form of the ligand. A broad peak with little intensity can also be observed arising from the ligand in the DNA-bound state. Addition of another equivalent of ligand (2 eq. Ligand) shifts the equilibrium towards the binary complex. Thus, solely a broad NH peak of the ligand in the bound conformation is observed. The line broadening effect mainly arises from the shorter transverse relaxation time (T2) in the bound from (increased molecular weight). The complex formation seems to be saturated at 2 equivalents of ligand **3**, as the addition of another equivalent does not change the appearance of the proton spectrum.

Rankings of Active, Unknown Compounds Found by ROCS

Table 1: New MGB identified by shape-based screening are listed together with the ranks they achieved by using the respective query molecules indicated as column headers. Their highest ranks in the NCI databases are indicated by italic numbers in the columns of the respective query molecules. 1PQQ containing a cyclic polyamide MGB does not contribute to the highest ranks. Nine out of ten new MGBs would have been found by inspecting only the best-ranked 200 compounds instead of 400.

	121D	1FTD	1JTL	1PQQ	227D
Compound **1**	129	2195	*31*	*33*	40468
Compound **2**	215	5220	*35*	*72*	62028
Compound **3**	*145*	*83*	1383	7153	66526
Compound **4**	3634	19663	20108	16140	*50*
Compound **5**	601	8385	12276	16946	*103*
Compound **6**	*144*	*119*	419	3379	6437
Compound **7**	1282	6771	450	3498	*171*
Compound **8**	6568	*43*	1683	*277*	42030
Compound **9**	*77*	888	3085	1478	3145
Compound **10**	2030	2290	*376*	1817	17562

Insoluble Compounds Found by ROCS

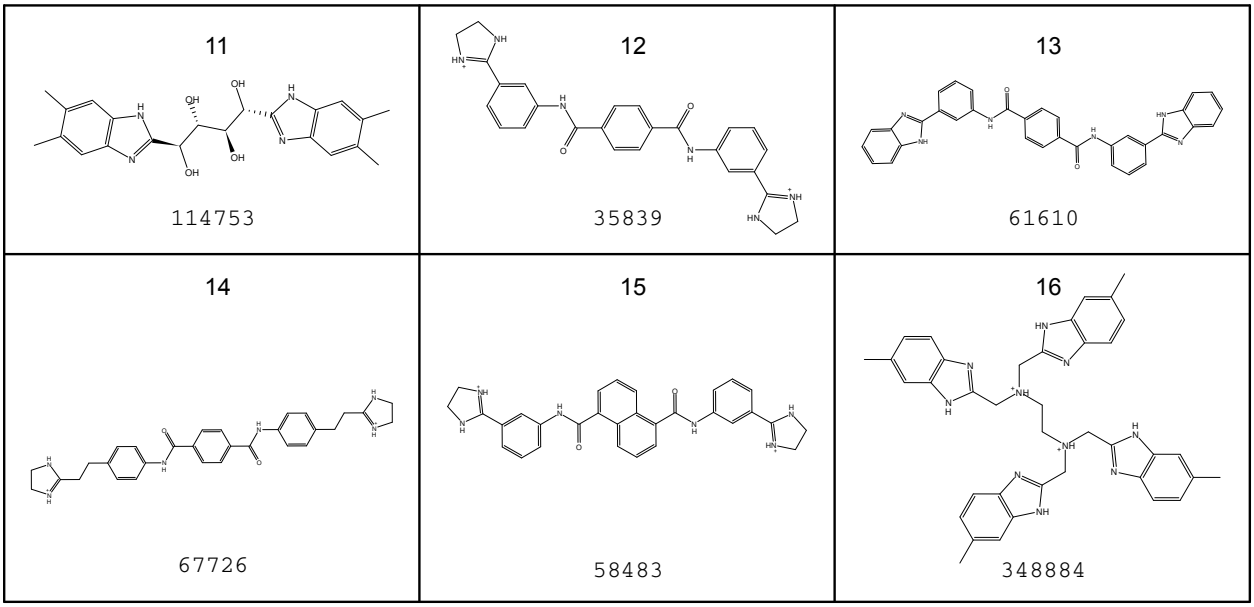


Figure 3: Compounds found by shape-based screening but not soluble under experimental conditions. Numbers at the bottom indicate NSC Codes (numbering scheme of the NCI database).

Table 2: Rankings for compounds found to be insoluble, the best rank values are marked italic.

	121D	1FTD	1JTL	1PQQ	227D
Compound **11**	8144	*107*	1570	7551	8552
Compound **12**	122	2069	*27*	*75*	4374
Compound **13**	2343	1615	*163*	*163*	90767
Compound **14**	950	314	*98*	*193*	19837
Compound **15**	1853	17132	*326*	*12*	15545
Compound **16**	119067	5578	39494	*73*	19837

Inactive Compounds Found by ROCS

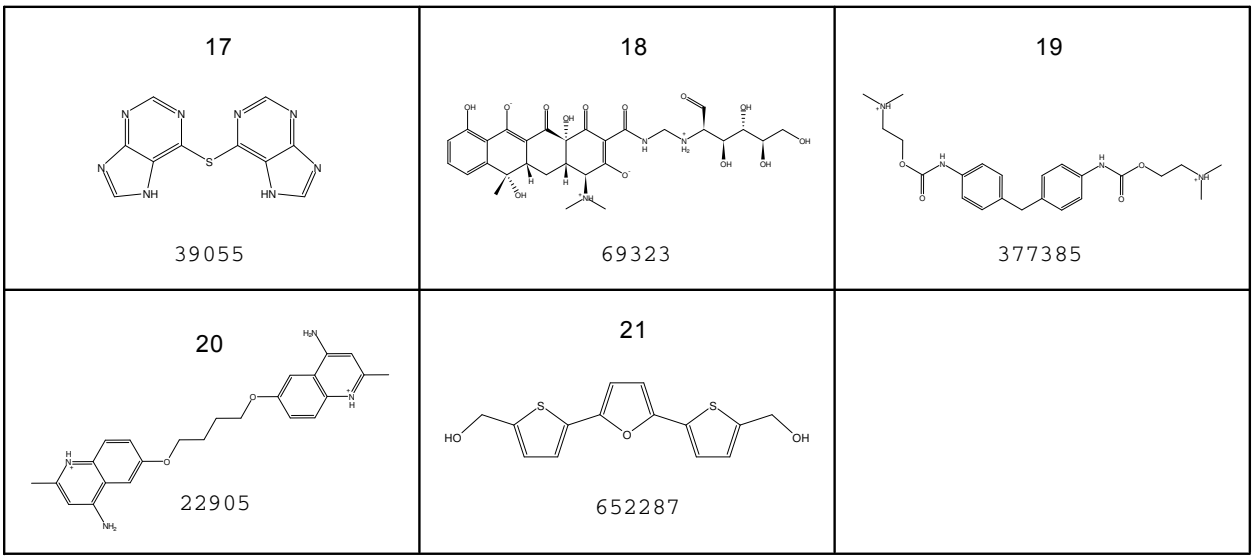


Figure 4: Compounds found by shape-based screening which showed DNA affinity larger than 100 μM (ligand **20**) or no affinity at all (all other ligands). Numbers at the bottom indicate NSC Codes (numbering scheme of the NCI database).

Table 3: Rankings for inactive compounds, the best rank values are marked italic. Results of one further query are reported here (1D64, the ligand is pentamidine) as it did not help to find additional active MGBs but was responsible for testing ligand 3h.

	121D	1FTD	1JTL	1PQQ	227D	1D64
Compound **17**	19742	*221*	37902	39225	44643	83965
Compound **18**	39688	13386	12206	*291*	61968	16585
Compound **19**	*66*	2764	426	25747	20613	9893
Compound **20**	18854	3437	3700	11365	18847	*234*
Compound **21**	12119	30881	28571	76032	*24*	19659

Active Compounds Found by ROCS - Known MGBs

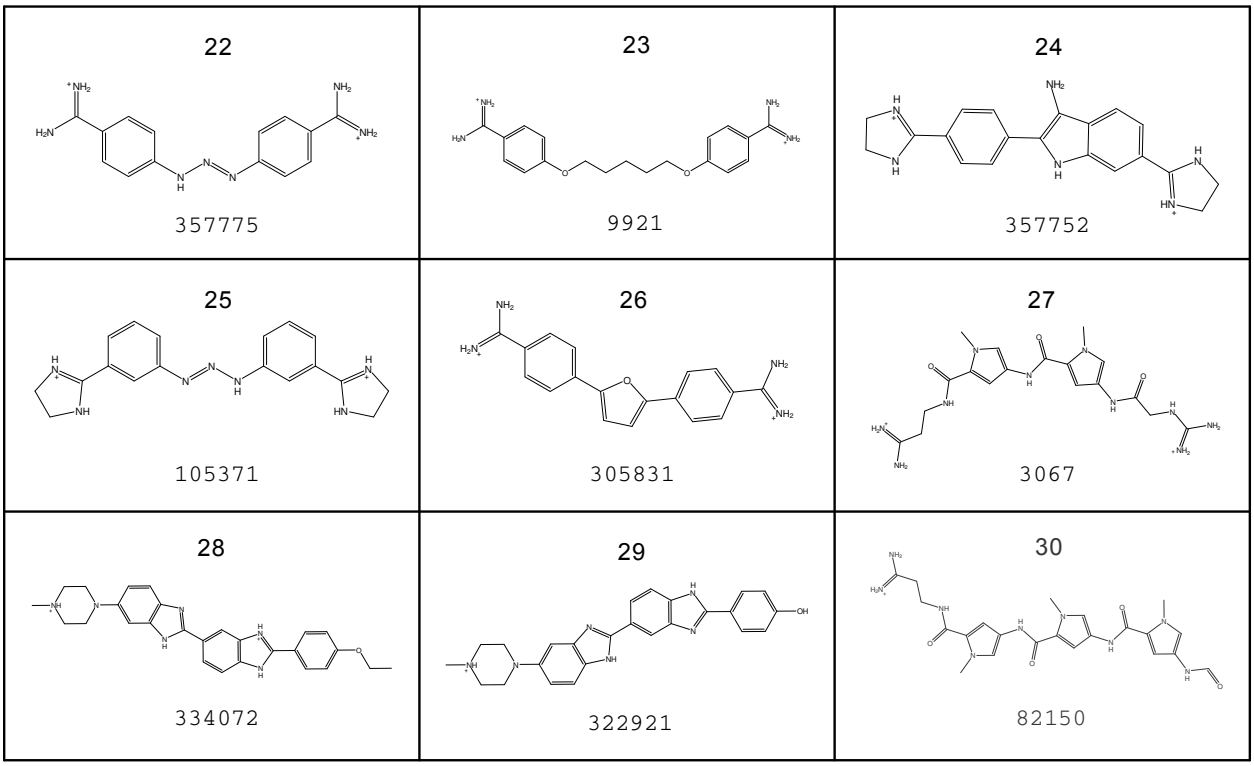


Figure 5: These compounds were found in the NCI database by ROCS shape-based screening. They are known MGBs well documented in the literature. Numbers at the bottom indicate NSC Codes (numbering scheme of the NCI database).

Table 4: Rankings for active compounds known to be MGBs, the best rank values are marked italic. Rank one means that a ligand found in the NCI database is identical with a query ligand.

	121D	1FTD	1JTL	1PQQ	227D
Compound **22**	5912	27705	15156	69839	*9*
Compound **23**	430	5650	1131	10144	*156*
Compound **24**	1224	4059	3757	53369	*82*
Compound **25**	2123	2937	7264	13390	*21*
Compound **26**	7553	20461	60586	91704	*1*
Compound **27**	*1*	821	*3*	*76*	3682
Compound **30**	*9*	931	*1*	*3*	63954
Compound **28**	*127*	*3*	803	6948	3391
Compound **29**	*123*	*6*	540	5301	1583

Compounds Not Found by ROCS But Used in Experiments

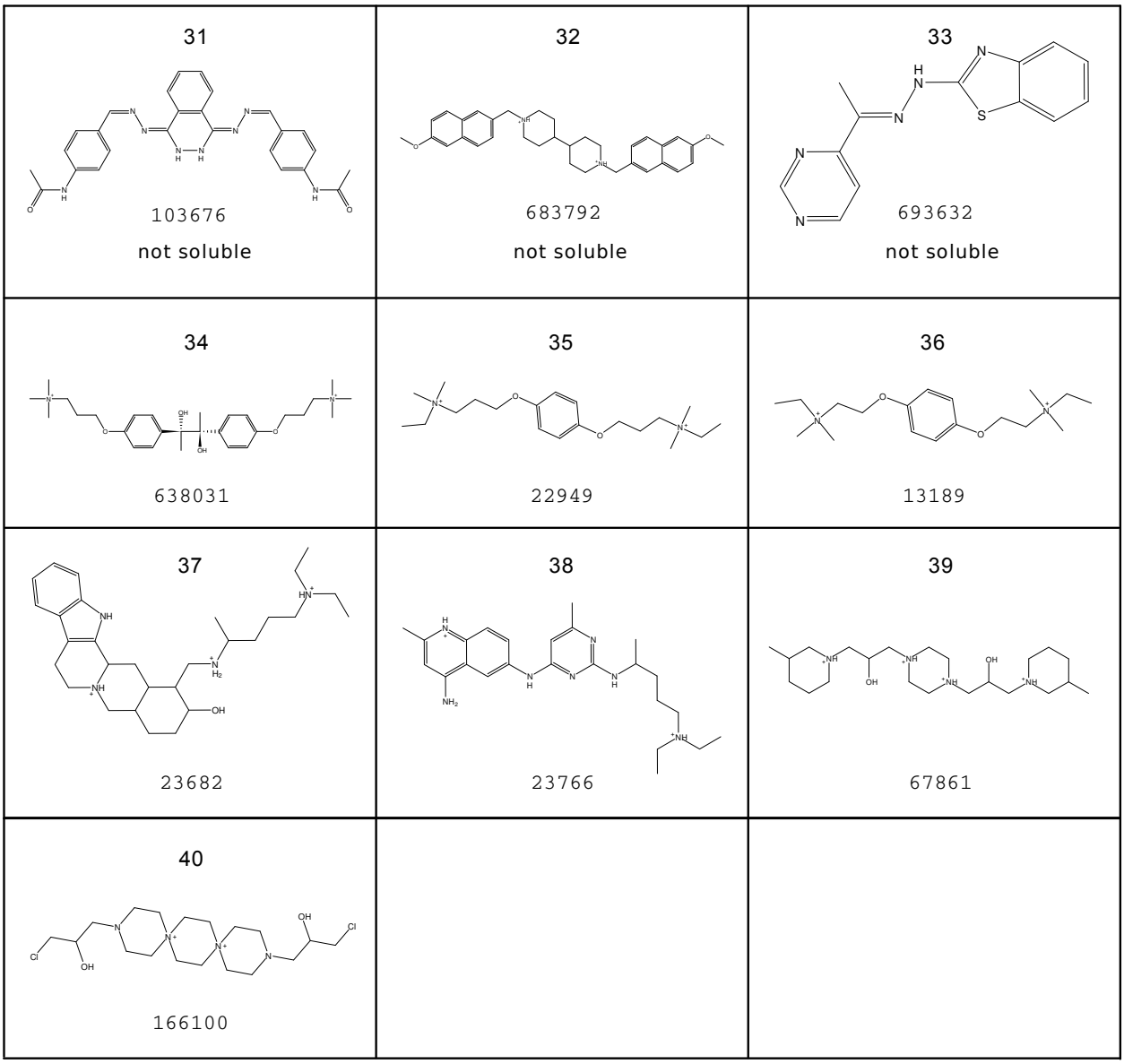


Figure 6: These compounds were selected without virtual screening. They all are inactive or have an affinity larger than 100 μM (ligands **37** and **38**). Compounds **31**, **32**, and **33** are just not soluble under experimental conditions. All compounds were arbitrarily selected because of their elongated structures. Numbers at the bottom indicate NSC Codes (numbering scheme of the NCI database).

MGBs and FDA-Approved Compounds Ranked in Large Databases

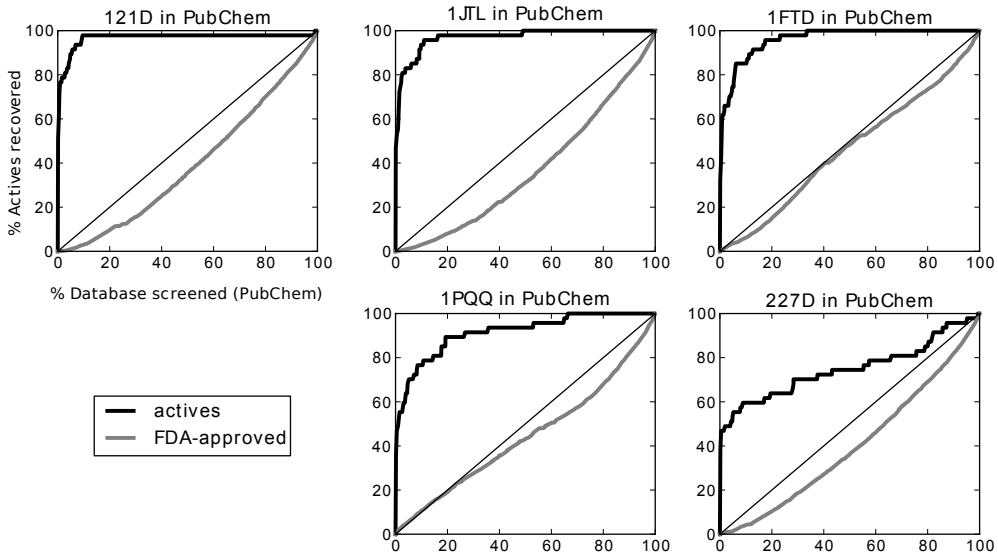


Figure 7: Ranking of active MGBs by ROCS shape-based screening in the PubChem database (121,207 entries) shows again high enrichment, as already seen for other databases. The same screening approach leads to a derichment of FDA-approved drugs in comparison with the PubChem database.

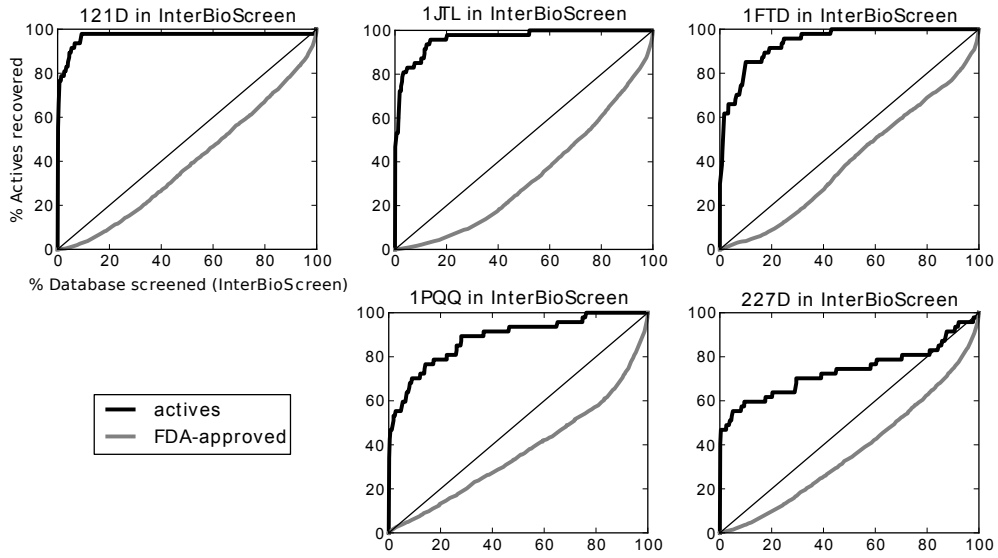


Figure 8: Ranking of active MGBs by ROCS shape-based screening in the InterBioScreen database (569,699 entries) shows again high enrichment, as already seen for other databases. The same screening approach leads to a derichment of FDA-approved drugs in comparison with the InterBioScreen database.

MOE 2D-Descriptors Tested

Following MOE 2D-Descriptors were tested in order to efficiently enrich minor groove binders in large data sets without applying more time-consuming shape-based screening. No single descriptor nor combination was found to be able to replace minor groove binder shape description by ROCS shape-based screening. Descriptions for the metrics are extracted from the MOE manual (MOE version 2009.10).

a_aro Number of aromatic atoms.

a_don Number of hydrogen bond donor atoms (not counting basic atoms but counting atoms that are both hydrogen bond donors and acceptors such as -OH).

a_nN Number of nitrogen atoms.

b_1rotN Number of rotatable single bonds. Conjugated single bonds are not included (e.g. ester and peptide bonds).

b_1rotR Fraction of rotatable single bonds: b_1rotN divided by b_heavy.

b_rotR Fraction of rotatable bonds: b_rotN divided by b_heavy.

balabanJ Balaban's connectivity topological index.[1]

b_ar Number of aromatic bonds.

bpol Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens) with polarizabilities taken from Szytula and Leciejewicz.[2]

chi0 Atomic connectivity index (order 0) from Hall and Kier.[3,4] This is calculated as the sum of 1/sqrt(d_i) over all heavy atoms i with $d_i > 0$.

chi0_C Carbon connectivity index (order 0). This is calculated as the sum of 1/sqrt(d_i) over all carbon atoms i with $d_i > 0$.

chi0v Atomic valence connectivity index (order 0) from Hall and Kier.[3,4] This is calculated as the sum of 1/sqrt(v_i) over all heavy atoms i with $v_i > 0$.

chi0v_C Carbon valence connectivity index (order 0). This is calculated as the sum of 1/sqrt(v_i) over all carbon atoms i with $v_i > 0$.

chi1 Atomic connectivity index (order 1) from Hall and Kier.[3,4] This is calculated as the sum of 1/sqrt($d_i d_j$) over all bonds between heavy atoms i and j where i < j.

chi1_C Carbon connectivity index (order 1). This is calculated as the sum of 1/sqrt($d_i d_j$) over all bonds between carbon atoms i and j where i < j.

chi1v Atomic valence connectivity index (order 1) from Hall and Kier.[3,4] This is calculated as the sum of 1/sqrt($v_i v_j$) over all bonds between heavy atoms i and j where i < j.

chi1v_C Carbon valence connectivity index (order 1). This is calculated as the sum of 1/sqrt($v_i v_j$) over all bonds between carbon atoms i and j where i < j.

Kier1 First kappa shape index.[3]

Kier2 Second kappa shape index.[3]

Kier3 Third kappa shape index.[3]

KierA1 First alpha modified shape index.[3]

KierA2 Second alpha modified shape index.[3]

KierA3 Third alpha modified shape index.[3]

KierFlex Kier molecular flexibility index: (KierA1) (KierA2) / n where n denotes the number of atoms in the hydrogen suppressed graph.[3]

zagreb Zagreb index: the sum of d_i^2 over all heavy atoms i.

lip_don The number of OH and NH atoms.

opr_nrot The number of rotatable bonds according to Oprea.[5]

PEOE_PC+ Total positive partial charge: the sum of the positive q_i with q_i denoting the partial charge of atom i.

PEOE_RPC+ Relative positive partial charge: the largest positive q_i divided by the sum of the positive q_i.

petitjean Value of (diameter - radius) / diameter.

petitjeanSC Petitjean graph Shape Coefficient as defined by Petitjean:[6] (diameter - radius) / radius.

radius If r_i is the largest matrix entry in row i of the distance matrix D, then the radius is defined as the smallest of the r_i.[6]

VAdjMa Vertex adjacency information (magnitude): 1 + log2 m where m is the number of heavy-heavy bonds. If m is zero, then zero is returned.

VDistEq If m is the sum of the distance matrix entries then VdistEq is defined to be the sum of $\log^2$ m - $p_i \log^2 p_i$ / m where p_i is the number of distance matrix entries equal to i.

VDistMa If m is the sum of the distance matrix entries then VDistMa is defined to be the sum of $\log^2$ m - $D_{ij} \log^2 D_{ij}$ / m over all i and j.

Weight Molecular weight (including implicit hydrogens) in atomic mass units with atomic weights taken from Szytula and Leciejewicz.[2]

weinerPath Wiener path number: half the sum of all the distance matrix entries as defined in Balaban [7] and Wiener.[8]

weinerPol Wiener polarity number: half the sum of all the distance matrix entries with a value of 3 as defined in Balaban.[7]

References

(1) Balaban, A. T. *Chemical Physics Letters* **1982**, *89*, 399–404.

(2) Szytula, A.; Leciejewicz, J. *Handbook of crystal structures and magnetic properties of rare earth intermetallics*; CRC, 1994.

(3) Hall, L.; Kier, L. In *Reviews of Computational Chemistry*; Boyd, D., Lipkowitz, K., Eds.; VCH Publishers, New York, NY (USA), 1991; Vol. 2, Chapter The molecular connectivity chi indices and kappa shape indices in structure-property modeling, pp 367–422.

(4) Kier, L.; Hall, L. *Eur. J. Med. Chem* **1977**, *12*, 334.

(5) Oprea, T. I. *J. Comp. Aid. Mol. Des.* **2000**, *14*, 251–264.

(6) Petitjean, M. *Journal of Chemical Information and Computer Sciences* **1992**, *32*, 331–337.

(7) Balaban, A. *Theor. Chim. Acta* **1979**, *53*, 355–375.

(8) Wiener, H. *Journal of the American Chemical Society* **1947**, *69*, 17–20.