

SUPPLEMENTARY INFORMATION

METHOD

Selection of Probe molecules. There have been two key considerations in new probe molecules design: probe molecules should (i) be small in size so that within an affordable MD simulation time (of the order of nanoseconds) they would be able to sample a sufficiently large fraction of the conformational space and efficiently explore the entire protein surface, and (ii) exhibit a range of drug-like physicochemical properties. To satisfy the first criterion we considered probes that share the same topology with isopropanol, i.e. four heavy atoms three of which are bonded to a central atom. To satisfy the second criterion, we analyzed the frequency of occurrence of small organic fragments as substructures in drugs. Drug molecule data were obtained from DrugBank version 3¹ and were analyzed using OpenBabel.²

First we investigated the different atom type sets and partial charge distributions for isopropanol to understand the dependence of the results on small variations in parameters (associated with different atom types) and partial charges. General rules deduced from this (see *Results*) were used to define the atom types and partial charge distributions of other probes. We have developed parameters for acetamide, acetate, and protonated isopropylamine (Table S1). These molecules are defined in CHARMM format.³ Atom types and partial charges used in probe definitions are adopted from amino acid residue definitions in CHARMM force field (Table S1).^{3,4} As for water molecules, TIP3 model was used.⁵

Designing new probes with the isopropanol topology had also an algorithmic advantage. In the proposed methodology, a probe's location in the simulation is tracked by the position of its central atom. This gives a good representation of the probe distribution in the system. Additionally, we evaluated benzene as a potential probe, but it had shortcomings as none of its heavy-atoms atoms described its center of mass. Yet, despite the absence of an aromatic probe molecule, the current set of probe molecules has enabled estimating maximal affinities for a diverse set of binding sites.

Probe and water mixture. The probe-water mixture used in system setup was composed of 6,860 water and 343 isopropanol molecules (Figure 1C), which gives a ratio of 20 water molecules per probe molecule. The mixture was a cubic box in which isopropanols were evenly distributed. The edges of the box were 62.36 Å long. Mixtures containing different probe compositions were obtained using PSFGEN⁶ by 'mutating' isopropanols to other probe types until the desired composition was reached (mutate is a PSFGEN command to alter the type of a residue).

Reference simulation and expected occupancy. We performed a 20 ns simulation of a reference mixture to calculate *expected occupancy* (Figure 1C). The reference mixture contained 207 isopropanol (60%), 34 isobutane (10%), 34 isopropylamine (10%), 34 acetate (10%), and 34 acetamide (10%) molecules, in addition to 6,860 water molecules. In the simulation, the probe molecules were restrained at their initial central carbon positions using a harmonic potential with force constant of 1 kcal/mol (applied to only central carbon atoms). Hence, probes were completely solvated and free to rotate during the simulation. Expected occupancy used in

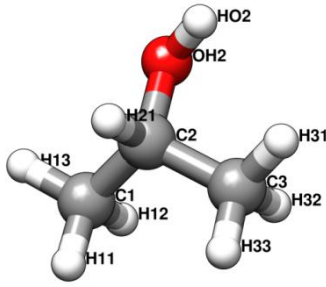
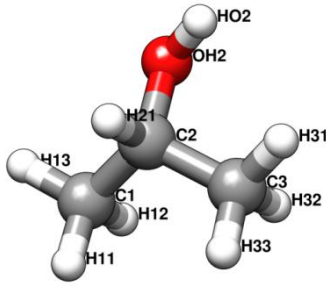
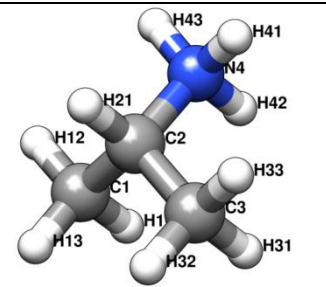
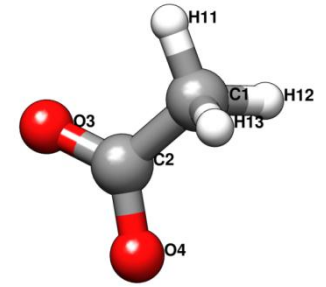
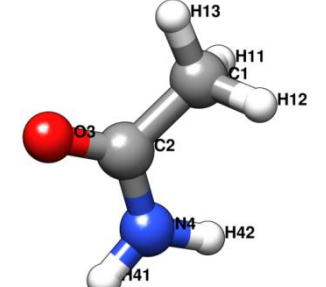
Equation 1 was calculated using the average volume of this *reference simulation* (240,930 Å³) and the relation $n_0 = V_0 \frac{N_{probes}}{V_{reference}}$. The expected occupancy in one cubic Ångstrom volume, for example, is $1 \text{ Å}^3 \frac{343}{240,930 \text{ Å}^3} = 14.2 \times 10^{-4}$. We also calculated n_0 using unconstrained probe-water simulations. The difference between two values was less than 1% of the significant figures in n_0 .

Effective probe radius. Probe interaction spots were considered as spheres. An effective radius, obtained from volume calculations (Table S3), was assigned to each probe molecule to enable calculating the volume of interaction spots.

System preparation. Protein structures were obtained from the Protein Data Bank.⁷ Appropriate histidine protonation states were determined using MOE.⁸ Missing side-chains were modeled using PSFGEN.⁶ All functional and structural cations were retained. All test case proteins were immersed in a solvent mixture box with padding distance of at least 6 Å along each direction from the protein. Solvent mixtures in all systems had a fixed ratio of 20 water molecules per probe. Different mole fractions for the probes were investigated as discussed in the main text. Solvated system coordinates were prepared using the VMD⁶ plugins Solvate, Autoionize, and PSFGEN.

MD simulations. To achieve an even distribution of probe molecules in the system, a simulated annealing protocol was implemented prior to the productive run. Simulated annealing was followed by equilibration of the system, for a total of 0.9 ns of simulation time. Longer annealing and equilibration times were also tested, but a total of 0.9 ns was found to be adequate for reproducible results (see Table S2).

Table S1. Isopropanol (IPRO) structure, CHARMM atom types, and partial charges

	Final Version [*]		CHARMM General ^{9**}		Seco et al. ¹⁰ and CHARMM ^{4 ***}		
	Atom Name	Atom Type	Partial Charge	Atom Type	Partial Charge	Atom Type	Partial Charge
	OH2	OH1	-0.660	OG311	-0.650	OH1	-0.612
	HO2	H	0.430	HGP1	0.420	H	0.373
	C2	CT1	0.181	CG311	0.140	CT1	0.309
	H21	HA	0.049	HGA1	0.090	HA	-0.012
	C1, C3	CT3	-0.147	CG331	-0.270	CT3	-0.176
	H11, H12, H13, H31, H32, H33	HA	0.049	HGA3	0.090	HA	0.049
	N4	NH3	-0.300				
	H41, H42, H43	HC	0.333				
	C2	CT1	0.252				
	H21	HA	0.049				
	C1, C3	CT3	-0.147				
	H11, H12, H13, H31, H32, H33	HA	0.049				
	C2	CC	0.520				
	O3, O4	OC	-0.760				
	C1	CT3	-0.147				
	H11, H12, H13	HA	0.049				
	C1	CT3	-0.147				
	H11, H12, H13	HA	0.049				
	C2	CC	0.550				
	O3	O	-0.550				
	N4	NH2	-0.620				
	H41	H	0.320				
	H42	H	0.300				

* All atom types and partial charges of hydroxyl group atoms are based on threonine residue definition in CHARMM force field.⁴ Partial charges of hydrocarbons in this and following probe molecules are based on partial charge distribution proposed by Seco et al.¹⁰

** Atom types and partial charges are based on CHARMM general force field.⁹

*** Atom types are based on threonine residue definition in CHARMM force field.⁴ Partial charges are from Seco et al.¹⁰

* Atom types and partial charges are based on lysine residue definition in CHARMM force field.⁴

* Atom types and partial charges are based on aspartic acid residue definition in CHARMM force field.⁴

* Atom types and partial charges are based on asparagine residue definition in CHARMM force field.⁴

Table S2. Description of simulation parameters and conditions, and predicted druggabilities/affinities*

Target (PDB id)	Run #	S1 (ps)	S2 (ps)	S3 (ps)	S4 (ps)	S5 (ps)	S6 (ps)	Sim. (ms)	Probe mole fractions	Predicted affinity	# of spots	# of atoms	# of CPUs	# of days
MDM2¹¹ PDB: 1YCR Chain: A	1-1	20 ^V	40 ^V	28	400	12	400	40	1 IPRO	0.5 nM	7	7,358	12	4.5
	1-2	40 ^P	80 ^P	60	600	20	600	40	1 IPRO	1.0 nM	7	7,934	12	4.2
	1-3	50 ^V	50 ^V	40	200	40	400	32	1 IPRO	0.4 nM	7	8,870	16	2.0
	1-1, 1-2, & 1-3									0.2 nM	7			
	1-4	20 ^V	40 ^V	28	400	12	400	40	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	1.3 nM	7	7,880	12	4.2
	1-5	40 ^P	80 ^P	60	600	20	600	40	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	0.3 nM	7	7,880	12	4.2
	1-6	20 ^V	40 ^V	28	400	12	400	32	0.4 IPRO, 0.2 ACET, 0.2 IPAM, 0.2 ACAM	2.0 nM	7	7,472	12	3.4
	1-4, 1-5, & 1.6									0.6 nM	7			
	1-7	50 ^V	50 ^V	40	200	40	400	32	1 IPRO	0.1 nM	7	9,014	8	3.4
	1-8	50 ^V	50 ^V	40	200	40	400	32	1 IPRO	1.0 nM	7	9,014	16	2.0
	1-9	50 ^V	50 ^V	40	200	40	400	32	1 IPRO	0.05 nM	7	8,870	10	3.1
1-10	20 ^V	40 ^V	-	-	-	-	-	40	Probe free simulation.	-	-	6,563	12	3.8
1-11	20 ^V	40 ^V	-	-	-	-	-	40	Probe free simulation.	-	-	6,563	12	3.8
MDM2¹² PDB: 1Z1M Model: 2	2-1	20 ^V	40 ^V	28	400	12	400	40	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	0.4 nM	7	28,332	24	7.4
	2-2	20 ^V	40 ^V	28	400	12	400	40	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	1.0 nM	7	28,332	4+GPU	6.1
	2-1 & 2-2									1.8 nM				
	2-3	20 ^V	40 ^V	-	-	-	-	40	Probe free simulation.			26,127	24	4.6
PTP1B¹³ PDB: 1PH0 Chain: A	3-1	20 ^V	40 ^V	28	400	12	400	40	1 IPRO	nd	-	21,008	96	1.6
										2.8 μM	7			
	3-2	20 ^V	40 ^V	28	400	12	400	40	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	0.3 nM	7	21,014	48	2.5
										17.6 μM	6			
	3-3	20 ^V	40 ^V	28	400	12	400	32	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	0.9 nM	7	20,876	16	6.8
										9.5 μM	7			
	3-2 & 3-3									1.3 nM	7			
										8.5 μM	7			
	3-4	20 ^V	40 ^V	-	-	-	-	40	Probe free simulation.	-	-	19,484	12	9.3
	3-5	20 ^V	40 ^V	-	-	-	-	40	Probe free simulation.	-	-	19,484	12	9.6
LFA-1¹⁴ PDB: 1ZOP Chain: A	4-1	20 ^V	40 ^V	28	400	12	400	40	1 IPRO	0.5 nM	7	13,337	12	6.9
	4-2	20 ^V	40 ^V	28	400	12	400	40	1 IPRO	0.8 nM	7	14,057	12	7.4
	4-1 & 4-2									0.4 nM	7			
	4-3	20 ^V	40 ^V	28	400	12	400	40	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	0.5 nM	7	13,679	12	7.2
	4-4	20 ^V	40 ^V	28	400	12	400	40	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	0.03 nM	7	14,393	24	4.4
	4-3 & 4-4									0.08 nM	7			
	4-5	20 ^V	40 ^V	-	-	-	-	40	Probe free simulation.	-	-	12,242	12	6.5
4-6	20 ^V	40 ^V	-	-	-	-	40	Probe free simulation.	-	-	12,242	12	6.6	
Eg5¹⁵ PDB: 1II6 Chain: A	5-1	50 ^V	50 ^V	40	200	40	400	32	1 IPRO	23 nM	7	35,822	10	9.5
	5-2	20 ^V	40 ^V	28	400	12	400	40	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	0.3 nM	7	31,214	60	2.8
	5-3	20 ^V	40 ^V	-	-	-	-	40	Probe free simulation.	-	-	29,921	80	2.5
p38¹⁶ PDB: 1P38 Chain: A	6-1	40 ^P	80 ^P	60	600	20	600	40	1 IPRO	2 nM	8	27,411	120	1.8
	6-2	40 ^P	80 ^P	60	600	20	600	40	1 IPRO	1 nM	8	28,563	140	1.8
	6-1 & 6-2									3.5 nM	8			
	6-3	40 ^P	80 ^P	60	600	20	600	40	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	0.12 nM	8	27,801	12	13
	6-4	40 ^P	80 ^P	60	600	20	600	40	0.7 IPRO, 0.1 ACET, 0.1 IPAM, 0.1 ACAM	0.01 nM	8	28,515	140	1.7
	6-3 & 6-4									0.09 nM	8			
	6-5	20 ^V	40 ^V	-	-	-	-	40	Probe free simulation.	-	-	26,454	12	13
6-6	20 ^V	40 ^V	-	-	-	-	40	Probe free simulation.	-	-	26,454	12	13	

*PDB structure and chain identifiers are given in column 1. Systems were subject to 2,000 steps of minimization prior to equilibration. Columns S1-S6 refer to the duration of various steps of the equilibration. In the 1st step of equilibration, S1, the system temperature was raised from 100K to 300K. S2 was run at 300K. In S1 and S2, either the volume or the pressure of the system remained constant, as indicated by superscripts. In S3, the temperature was raised from 300K to 600K. S4 was run at 600K. In S5, the temperature was decreased to 300K. In steps S3, S4, and S5, the system volume remained constant. In S1-5, C^α atoms were restrained by a harmonic potential with a force constant of 1 kcal/mol/Å². In S6, the system was simulated at constant pressure (1 atm) and temperature (300K) without constraints. Column 9 lists the mole fraction of the probe molecules in each run, acronyms IPRO, IPAM, ACAM, and ACET being used for isopropanol, isopropylamine, acetamide, and acetate, respectively. Last three columns state the number of atoms in the systems, number of CPUs, and days to run the productive simulations. Sim 2-2 was run on a node with 4 CPUs and NVIDIA Tesla M2090 GPU.

Effective radii of probe molecules

We identified high affinity non-overlapping probe binding spots in occupancy grids by approximating the volume occupied by probes as a sphere. Effective radii for each probe were calculated from volumes obtained from the fragment-based molecular property calculation interface of Molinspiration¹⁷ (Table S3). We also compared these values to those calculated from bulk properties of pure liquids and radial distribution functions generated from simulations. Calculation of effective radii for solvent molecules requires consideration of packing densities, i.e. $V_{eff} = \frac{4}{3}\pi r_{eff}^3 \cong \frac{\rho 10^{24}}{MN_A} \rho^*$ where ρ^* is packing density, ρ is density, M is molar mass, N_A is Avogadro's number, 10^{24} is conversion factor from cm to Å. For water molecule, for example, calculation of the effective radius while omitting packing density, i.e. $\rho^* = 1$, results in 1.93 Å, which is considerably different from the widely accepted 1.4 Å radius established using molecular simulations. This value is matched when $\rho^* = 0.39$ is used. For alcohols and other organic solvents capable of hydrogen bonding, we found that packing densities range from 0.54 to 0.60.¹⁸ We calculated effective probe radii for $\rho^* = 0.58$ and compared to those we used in druggability analysis in Table S3. Radii from these two methods differ by less than 3%.

Table S3. Effective radii of probe molecules

Molecule	Computational ^a		Experimental ^b			
	Volume (Å ³)	Radius (Å)	Density (g/cm ³)	Molar mass (g/mol)	Volume (Å ³)	Radius (Å)
isopropanol	70.60	2.564	0.79	60.10	73.64	2.60
isopropylamine	73.87	2.603	0.72	60.12	80.20	2.68
acetate	56.20	2.376	1.05	59.05	54.21	2.35
acetamide	59.47	2.421	1.16	59.07	49.04	2.27

^a Computational values are calculated using Molinspiration software. ^b Experimental values are calculated using bulk properties of pure liquids with packing density correction, 0.58 for probe molecules.

Reference List

1. Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **2011**, *39*, D1035-D1041.
2. O'Boyle, N. M.; Morley, C.; Hutchison, G. R. Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J* **2008**, *2*, 5.
3. Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J Comput. Chem.* **2009**, *30*, 1545-1614.

4. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187-217.
5. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem. Phys.* **1985**, *79*, 926.
6. Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J Mol. Graph.* **1996**, *14*, 33-38.
7. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.
8. Molecular Operating Environment (MOE). <http://www.chemcomp.com/> . 2009. Chemical Computing Group Inc.
9. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput. Chem.* **2010**, *31*, 671-690.
10. Seco, J.; Luque, F. J.; Barril, X. Binding site detection and druggability index from first principles. *J Med. Chem.* **2009**, *52*, 2363-2371.
11. Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **1996**, *274*, 948-953.
12. Uhrinova, S.; Uhrin, D.; Powers, H.; Watt, K.; Zheleva, D.; Fischer, P.; McInnes, C.; Barlow, P. N. Structure of free MDM2 N-terminal domain reveals conformational adjustments that accompany p53-binding. *J Mol. Biol* **2005**, *350*, 587-598.
13. Liu, G.; Xin, Z.; Liang, H.; bad-Zapatero, C.; Hajduk, P. J.; Janowick, D. A.; Szczepankiewicz, B. G.; Pei, Z.; Hutchins, C. W.; Ballaron, S. J.; Stashko, M. A.; Lubben, T. H.; Berg, C. E.; Rondinone, C. M.; Trevillyan, J. M.; Jirousek, M. R. Selective protein tyrosine phosphatase 1B inhibitors: targeting the second phosphotyrosine binding site with non-carboxylic acid-containing ligands. *J Med. Chem.* **2003**, *46*, 3437-3440.
14. Qu, A.; Leahy, D. J. The role of the divalent cation in the structure of the I domain from the CD11a/CD18 integrin. *Structure.* **1996**, *4*, 931-942.
15. Turner, J.; Anderson, R.; Guo, J.; Beraud, C.; Fletterick, R.; Sakowicz, R. Crystal structure of the mitotic spindle kinesin Eg5 reveals a novel conformation of the neck-linker. *J Biol. Chem.* **2001**, *276*, 25496-25502.
16. Wang, Z.; Harkins, P. C.; Ulevitch, R. J.; Han, J.; Cobb, M. H.; Goldsmith, E. J. The structure of mitogen-activated protein kinase p38 at 2.1-Å resolution. *Proc. Natl. Acad. Sci. U. S. A* **1997**, *94*, 2327-2332.
17. Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J Med. Chem* **2000**, *43*, 3714-3717.
18. Bondi, A. *Physical Properties of Molecular Crystals, Liquids, and Glasses*; John Wiley & Sons, Inc.: New York, 1968.