

Structure-based maximal affinity model predicts small-molecule druggability

Alan C Cheng¹⁻³, Ryan G Coleman¹, Kathleen T Smyth², Qing Cao¹, Patricia Soulard², Daniel R Caffrey¹, Anna C Salzberg¹ & Enoch S Huang¹

Lead generation is a major hurdle in small-molecule drug discovery, with an estimated 60% of projects failing from lack of lead matter or difficulty in optimizing leads for drug-like properties. It would be valuable to identify these less-druggable targets before incurring substantial expenditure and effort. Here we show that a model-based approach using basic biophysical principles yields good prediction of druggability based solely on the crystal structure of the target binding site. We quantitatively estimate the maximal affinity achievable by a drug-like molecule, and we show that these calculated values correlate with drug discovery outcomes. We experimentally test two predictions using high-throughput screening of a diverse compound collection. The collective results highlight the utility of our approach as well as strategies for tackling difficult targets.

An estimated 60% of small molecule drug discovery projects fail in hit-to-lead because the biological target is found to be not 'druggable'¹. This is a combined attrition due to lack of lead matter from screening and difficulty in optimizing lead matter to yield reasonable drug leads. By some estimates, only 10% of genes in the human genome are druggable, and only 5% are both druggable and relevant to disease². These numbers are for 'druggability' defined as the likelihood of modulating a target by oral small-molecule drugs. Although other delivery approaches, such as injection or inhalation, have demonstrated success against targets that are considered difficult for orally bioavailable drugs, most industry efforts continue to focus on oral delivery owing to ease of dose administration and the potential to treat disease by modulating intracellular and central nervous system targets. High-throughput screening of new targets is costly; thus the ability to assess targets for their probability of success in advance of experimental screening can help reduce the high failure rate and cost of drug discovery.

The most common approach to estimating druggability is to classify targets by whether they belong to gene families known to be druggable, such as G protein-coupled receptors or kinases². However, gene families not known to be druggable have yielded novel targets through genomics

efforts, and not all members of a given gene family are equally druggable³. Here we show that a binding-free-energy model combined with recently discovered parameters for drug-like properties allows us to predict well the druggability of pharmaceutical targets based solely on binding-site structure. Conceptually, our model estimates the maximal achievable affinity for a binding pocket from a hit-to-lead optimization effort, given the constraint that the lead must have certain drug-like properties consistent with passive oral bioavailability.

The presence of a cleft or pocket on a biological target is, in general, necessary but not sufficient for modulation by drug-like small molecules. Although protein-protein interactions are generally considered undruggable, the recent case of the MDM2/p53 interaction inhibitors⁴ indicates that this is not always the case. We thus ventured to build a model for the druggability of a binding site based solely on the binding site's physiochemical properties. We started with two principles. The first is that orally bioavailable drugs have distinct physiochemical property ranges. Recent work has shown that passively absorbed oral drugs tend to fall within fixed ranges of size (≤ 500 Da) and polar surface area (≤ 140 Å²)⁵⁻⁷. In addition, such drugs rarely have more than one formal charge because only neutral molecules passively diffuse across membranes⁸. The

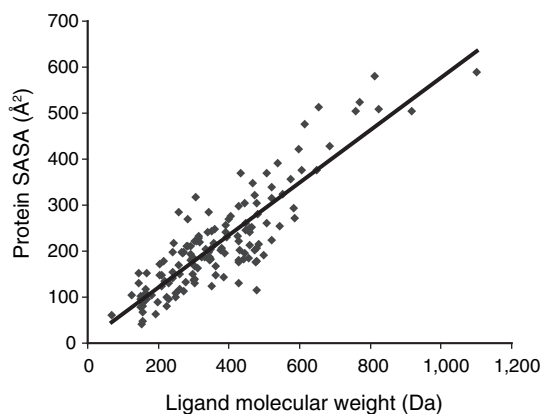


Figure 1 Ligand molecular weight correlates with protein-binding pocket surface area. A 500-Da ligand corresponds to about 300 Å² on the protein surface. Data measured from a diverse set of 305 protein-ligand co-crystal structures from CCDC/Astex³¹. The correlation is $r^2 = 0.77$. SASA, solvent-accessible surface area.

¹Departments of ¹Molecular Informatics and ²Biological Sciences, Research Technology Center, Pfizer Global Research & Development, Cambridge, Massachusetts 02139, USA. ³Present address: Amgen Inc., One Kendall Square Bldg. 1000, Cambridge, Massachusetts 02139, USA. Correspondence should be addressed to A.C.C. (alan.cheng@amgen.com).

Published online 8 January 2007; doi:10.1038/nbt1273

second is that the target binding site has a maximal achievable noncovalent binding affinity⁹ for drug-like compounds, and that this affinity can be calculated based on our current understanding of the molecular driving forces for binding.

We model the variation in maximal achievable binding energy for an optimized drug-like molecule as largely due to desolvation:

$$\Delta G_{MAP_{POD}} \approx \Delta G_{desolvation}^{target} + \Delta G_{desolvation}^{ligand} + \Delta G_{constant} \quad (1)$$

where $\Delta G_{desolvation}$ terms represent release of water upon binding from the target binding site and part of the ligand. Other binding-energy components include van der Waals and electrostatic interaction energies, and changes in translational, rotational and vibrational entropies, which we treat simply as a constant contribution, $\Delta G_{constant}$, to the maximal affinity predicted for a passively absorbed oral drug (MAP_{POD}). This is a reasonable approximation because we are considering binding of drug-like molecules similar in size (~500 Da) and number of charges (rarely more than one formal charge).

A simple model for desolvation is $\Delta G_{desolvation} = -\gamma \cdot A$ (ref. 10), where γ is related to solvent surface tension, and A is the relevant solvent-accessible surface area (SASA). It has been argued that polar surfaces make little contribution to the hydrophobic effect¹¹. Equation (1) can then be written:

$$\Delta G_{MAP_{POD}} \approx -\gamma(r)A_{nonpolar}^{target} - \gamma_{constant}A_{nonpolar}^{ligand} + \Delta G_{constant}, \text{ where } \gamma(r) = \frac{\gamma(\infty)}{1 - \frac{1.4}{r}} \quad (2)$$

For small ligands, a constant $\gamma_{constant} = 24 \text{ cal/mol/\AA}^2$ is well accepted¹⁰. For larger surfaces, a γ dependent on the curvature of the surface, r , is thought to be more appropriate^{12–14}. This curvature-dependent model for hydrophobic desolvation has been applied to convex surfaces such as alkanes¹², and here we apply it to concave surfaces found in ligand-binding pockets. The $\gamma(\infty)$ term in the model represents the hydrophobic desolvation of a flat surface^{12,15}. The model does not have an explicit polar desolvation term, but we implicitly use the argument that electrostatic interaction and desolvation energies for a charged group act in opposition, and the combination of the two generally provides only an insubstantial contribution¹⁶ to drug-like maximal affinity. Because $\Delta G_{MAP_{POD}}$ becomes more favorable with increasing ligand size, normalization for molecular size is necessary for comparisons. We found an

approximately linear correlation between ligand molecular weight and binding-site surface area, which allows us to normalize the total SASA of the protein pocket, A_{total}^{target} , to a drug-like size, $A_{druglike}^{target} = 300\text{\AA}^2$, representing a compound of ~500 Da in molecular weight (see Fig. 1 for details):

$$\Delta G_{MAP_{POD}} \approx -\gamma(r)A_{nonpolar}^{target} \frac{A_{druglike}^{target}}{A_{total}^{target}} + C \quad (3)$$

where the constant, C, includes the ligand desolvation component under the approximation that $A_{nonpolar}^{ligand}$ in equation (2) is a constant. In this work, we convert $\Delta G_{MAP_{POD}}$ values to K_d values more commonly used in drug discovery by using the equation, $K_d = \exp(\frac{\Delta G}{RT})$, where R is the gas constant and the temperature, T, is 298 K.

To implement the theory, we used computational geometry methods^{17,18} to represent the binding site and accurately calculate the necessary quantities of curvature and surface area from three-dimensional crystal structures. The model requires a structure of the bound-pocket conformation, but the observed ligand is not usually the maximal affinity ligand. In fact the ligands used in this study include bound peptides and substrate mimetics.

The parameters $\gamma(\infty)$ and C remain undefined. To determine a value for $\gamma(\infty)$, we initially set C to zero and looked for differentiation on a set of eight crystal structures (listed in Methods) representing targets that range in experimental druggability. We empirically found that values of $\gamma(\infty) = 45 \text{ cal/mol/\AA}^2$ and C = 0 largely discriminate druggable from undruggable targets. We note that the model itself is physically derived, and that the $\gamma(\infty)$ value we find is nearly identical to an experimentally measured value¹⁵, whereas the value of zero for C can be explained but requires further investigation (see comments in Supplementary Methods online).

We next applied the method to 63 structures representing 27 pharmaceutical targets (listed in Supplementary Table 1 online), including 23 targets with marketed drugs and four ‘undruggable’ targets that have been pursued extensively by multiple pharmaceutical companies with little success. We refer to the four targets as ‘undruggable’, although we recognize that it is always possible that a drug-like small molecule may eventually be discovered. We found that all ‘undruggable’ targets in our data set have calculated druggability scores of $K_d > 100 \text{ nM}$, but the ones for druggable targets span a wide range, from 0.005 to 2,000 nM. Although druggable targets tend to have more favorable MAP_{POD} values, this discrimination did not appear at first to be useful for making decisions on target prosecution. However, further study of the five druggable targets scoring in the ‘undruggable’ target range found that the known drugs for these targets are highly polar and not passively absorbed, and instead require administration as prodrugs or the use of active transporter mechanisms.

In fact, the major drugs for neuraminidase, inosine monophosphate dehydrogenase (IMPDH), angiotensin-converting enzyme 1 (ACE-1) and the nucleotide site of HIV reverse transcriptase (HIV RT) are all

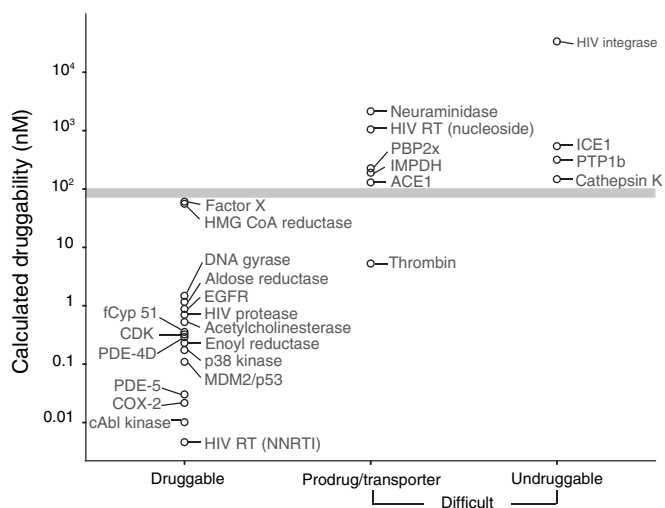


Figure 2 Calculated druggability for a set of 27 target binding sites. Known druggable protein targets are shown on the left vertical, whereas known difficult targets (prodrug and “undruggable”) are shown in the right verticals. Difficult and druggable target binding sites are effectively separated by the gray bar. The predicted druggability is the MAP_{pod} score calculated from the protein-ligand binding site structure. HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; EGFR, epidermal growth factor receptor kinase; CDK, cyclin-dependent kinase 2; PDE, phosphodiesterase; COX, cyclooxygenase; HIV RT, HIV reverse transcriptase; PBP2x, penicillin binding protein 2x; IMPDH, inosine monophosphate dehydrogenase; ACE-1, angiotensin-converting enzyme 1; ICE1, interleukin-1 β -converting enzyme 1; PTP1b, phosphotyrosine phosphatase 1B.

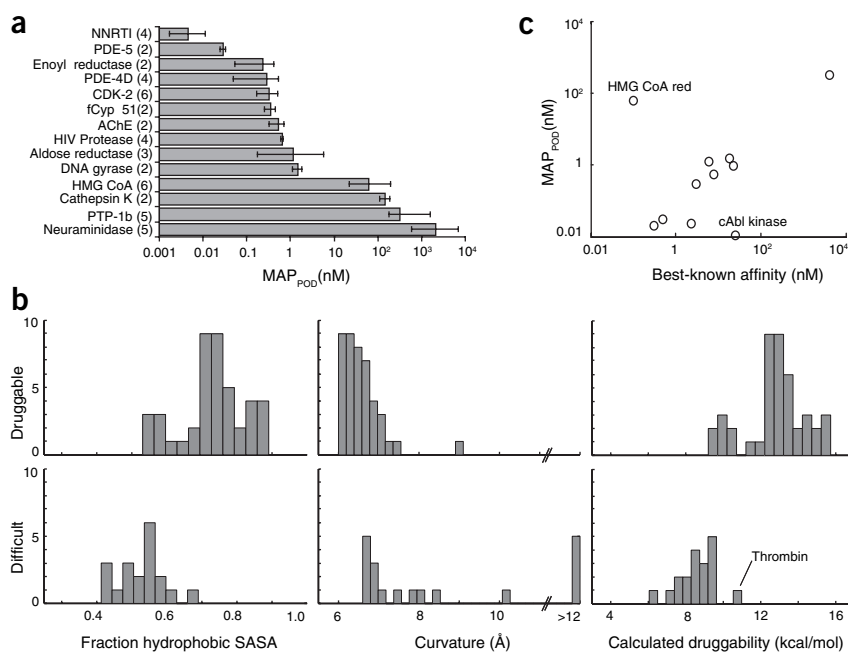


Figure 3 MAP_{POD} score comparisons.

(a) Variation in druggability predictions. Bars represent the median computed druggability of the target-binding sites. Range bars indicate the minimum and maximum predicted affinities (these are the full ranges, not the s.d.). Aldose reductase in the data set has a large variation in predictions, but the range does not affect its druggability classification. (b) Histograms of fraction hydrophobic SASA, pocket curvature and predicted druggability values for druggable and difficult (prodrug and 'undruggable') targets. Thrombin, which is poorly predicted by this method, is identified in the right hand histograms for predicted druggability. Histograms generated using Matlab 7.0.4 (Mathworks). (c) Correlation of predicted maximal achievable drug-like affinity with best-known affinity. The y-axis represents predicted maximal achievable drug-like affinity from our study, and the x-axis represents the lowest known affinity for a drug-like compound. Some error is expected because the data is composed of a mix of IC₅₀s, K_ds and K_s. Details of the data for lowest known affinity are available in **Supplementary Table 2** online.

administered as prodrugs. In the case of penicillin-binding protein (PBP2x), the β -lactam drugs are covalent inhibitors, and most if not all are also administered as prodrugs or are substrates of peptide transporters (see **Supplementary Table 1** online for details). Most of the prodrugs for the five targets are ester versions of the acid compound, where the acid is required for chelating an active-site Zn²⁺ or Mg²⁺, and in effect the prodrug moiety is used to mask the acid for increased oral absorption¹⁹. In the case of the neuraminidase drug oseltamivir (Tamiflu), the prodrug hides a carboxylic acid that forms multiple salt-bridges to three different arginine guanidinium groups²⁰, whereas with HIV RT nucleotide inhibitors, the prodrugs are nucleoside mimetics that are phosphorylated *in vivo*. Although our method classifies these targets in the undruggable range because it does not account for strong charge interactions or metal chelation, the chemical functionalities required for such interactions are also not conducive to passive oral bioavailability. Prodrugs substantially complicate drug development and are avoided when possible, and it is difficult to design for the use of active transporters because transporter selectivity is not well understood⁸.

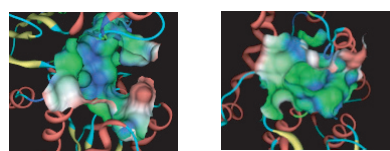
Thus, the five targets with marketed drugs predicted to have weaker than 100 nM MAP_{POD} affinities are in fact difficult targets with the current state of the art. Further dividing the druggable set into those requiring prodrug or active transport mechanisms and plotting the calculated maximal achievable drug-like affinities (MAP_{POD} values) of the 27 targets results in the plot in **Figure 2**. The gray horizontal bar indicates separation of the druggable and difficult druggability targets at 70–100 nM. Despite the simplicity of our biophysical model we are able to achieve surprisingly good discrimination. The one target incorrectly predicted here is thrombin, where the method classifies it in the druggable range but substantial effort has, to our knowledge, resulted in only prodrugs passing phase 2 clinical trials.

To estimate the error range of the calculations due to variation in binding mode and structural accuracy, we analyzed multiple cocrystal structures for 14 of the targets, selecting diverse cocrystallized ligands when possible (see **Fig. 3a**). The mean variation (maximum minus minimum values) in MAP_{POD} values is 0.9 kcal/mol, with a 99% statistical confidence interval of ± 0.5 kcal/mol calculated from a two-tailed *t*-test. This is equivalent to a tenfold or less variation in predicted K_d. Thus, analysis

of multiple cocrystal structures suggests that the druggability calculations are reasonably robust to small variations such as side-chain conformation changes. Binding sites undergo conformational change to differing extents to achieve their bound conformations, and this variation is not explicitly captured here. The current method is meant to address only the specified binding site and does not take into account the possibility of substantial flexibility or inhibition at allosteric pockets, both of which are important considerations in drug design^{21–23}.

The two variable properties in the model are curvature and hydrophobic surface area of the binding pocket. To determine how well each property discriminates between druggable and difficult, we generated histograms for each set of targets in **Figure 3b**. There is a clear trend toward higher fraction hydrophobic SASA and lower radius of curvature (binding site is more curved) for druggable targets, and there is greater overlap of the curvature distributions between the sets, suggesting that hydrophobicity by itself is a better discriminator. The combination of the two using the model, however, yields a substantially better discrimination than either property alone. Very recent data-mining studies using a large number of binding-site descriptors to discriminate small molecule-binding pockets from other pockets arrived at similar conclusions^{24,25}.

To address whether the calculated affinity values are quantitatively meaningful, we compared MAP_{POD} values for 11 targets with the affinities of the most potent known drug-like compounds (see **Supplementary Table 2** online for details). The identified experimental maximal affinity values are likely close to the true values given the substantial effort invested into small-molecule inhibitors for these targets. However, these values must be considered approximate due to the differences in assays used and the uncertainty over whether these are the true maximal affinity values. Nevertheless, **Figure 3c** suggests a reasonable linear correlation for the predicted and estimated experimental maximal-binding affinities. There are two outliers, HMG-CoA reductase and cAbl kinase. In the case of cAbl kinase, selectivity was likely more important than affinity for Gleevec, and a more potent (IC₅₀ < 1 nM) but less selective analog has recently entered clinical trials²⁶. In the case of HMG-CoA reductase, the predictions are three log orders greater than the best known drug affinity of 0.1 nM. This is likely due to underestimation of the contribution of electrostatic interactions, as statins have a conserved glutamyl group



	Fungal HSD	H-PGDS
MAP _{POD} K _D	240 nM	30 nM
Primary HTS hits	16	200
Compounds with IC ₅₀ ≤ 5 μM	2	33
Compounds with IC ₅₀ ≤ 1 μM	0	11

Figure 4 Predictions and screening results for two novel targets using a diverse set of 11,000 compounds. MAP_{POD} affinity scores for fungal homoserine dehydrogenase (HSD) and hematopoietic prostaglandin D synthase (H-PGDS) calculated using available crystal structures (PDB IDs: 1EBU and 1PD2, respectively). A set of compounds designed to be a diverse representation of known compound space were screened to find inhibitors of HSD and H-PGDS. High-throughput screening was performed at 20 μM compound concentration using duplicate measurements for each compound. Raw hits were selected based on hits meeting a 60% inhibition cutoff and significance of $P \leq 0.05$. Assay details available in the Methods section. The binding site surface of each target is shown and is based on the available crystal structures. Analytic Connolly surfaces are generated using MOE (version 2004.03; Chemical Computing Group) and is colored green for hydrophobic, blue for polar and red for solvent exposed.

that makes numerous hydrogen-bond and ion-pair interactions to the protein²⁷. Statin compounds, however, are all highly polar, with the top selling statins (atorvastatin, rosuvastatin and simvastatin) having polar surface areas between 150 and 174 Å², which is considerably higher than the drug-like range of <140 Å². Some first generation statins are delivered as prodrugs, and the mechanism of statin drug action requires high first-pass effect but does not require high bioavailability because the liver is the target organ²⁸. Thus despite the exceptions and the approximate nature of this maximal affinity study, the good correlation is very encouraging.

To test our method in a forward prediction, we used high-throughput screening (HTS) as a measure of druggability and compared this to MAP_{POD} druggability predictions for two novel targets. Fungal homoserine dehydrogenase (HSD) is a key enzyme in the essential aspartyl amino acid biosynthetic pathway of human fungal pathogens, and thus is a target for antifungal therapy. Hematopoietic prostaglandin D synthase (H-PGDS), is responsible for the production of prostaglandin D₂ (PGD₂), and thus is a key mediator of inflammatory responses. These targets represent particularly good test cases because no data were available on the druggability of the targets before the experiment, and both targets belong to gene families whose druggability is unknown.

We calculated MAP_{POD} affinities for both targets using cocrystal structures available in the Protein Databank (PDB), and found scores of $K_d = 240$ nM for HSD (PDB ID: 1EBU) and $K_d = 30$ nM for H-PGDS (PDB ID: 1PD2), predicting that HSD would be a difficult target whereas H-PGDS would be druggable. To assess druggability experimentally, we performed two HTS experiments, both using the same 11,000 compound 'diversity' collection designed to represent drug-like chemical space. We found 16 and 200 raw hits for HSD and H-PGDS, respectively. Moreover, the screens resulted in only two hits with IC₅₀s under 5 μM for HSD, whereas they identified 33 hits under 5 μM for H-PGDS (Fig. 4). Considerable follow-up efforts at Pfizer did not result in finding of a drug-like lead for HSD, whereas we were able to identify 11 sub-μM leads for H-PGDS from a relatively small HTS. These results lend support to our a priori druggability predictions and suggest how the method can affect prioritization of targets for drug discovery efforts. A substantial follow-on effort is required to determine whether H-PGDS is truly druggable based on the metric of successful advancement through clinical trials, and the success of targets is additionally influenced by aspects such as efficacy, safety and commercial attractiveness, which we cannot account for in a physical model. The identification of multiple potent drug-like hits from the diversity screening set nonetheless supports the predictivity and practical utility of our model.

The druggability model we have developed is consistent with both quantitative studies and qualitative intuition that favorable drug binding is largely driven by the hydrophobic effect²³, and we have shown that the model is useful in predicting the screening and lead-optimization outcome for targets with unknown experimental druggability.

We specifically attempted to model nonpolar desolvation here, and this appears to be the major contribution to variation in maximal binding affinity. Implementing the model necessitated use of recently developed computational geometry algorithms as opposed to standard available software because of the need to represent the binding site in a precise and physically reasonable way. For instance, using standard available software to perform a simple additive summation of the surface areas of atoms in a binding site results in overestimation of binding surface areas by about 40%, largely from contributions from the 'lip' of the pocket, which contributes minimally to binding affinity (A.C.C., R.G.C., unpublished data). It is conceivable, however, that available descriptors for surface areas and curvature can be combined and weighted to yield similar results. In fact, Hajduk *et al.*²⁵ recently used linear regression to fit NMR screening data (a reasonable measure of druggability) to commercially available structure-based descriptors of binding pockets. Although the regression equation contains a large number of terms, the dominant terms are exceptionally consistent with our findings, and include the pocket compactness (which correlates with curvature) and the nonpolar and total surface areas.

Drug discovery is extremely complex and difficult, and drug action is much more than binding affinity⁸. Nevertheless, the intrinsic biophysical maximal affinity modeled here appears to be a good estimate of the difficulty of drug discovery efforts, and has implications in target selection and setting of execution strategy. A priori druggability assessment can be used in combination with target validation and feasibility assessments in selecting targets. Marginally druggable targets may be planned for by, for instance, prioritizing structure-based drug design resources for guiding hit-to-lead campaigns or screening at higher compound concentrations. Our results also highlight strategies for difficult druggability targets, including consideration of covalent adduct, metal chelation, prodrug, active transport and allosteric approaches. Although we have focused on protein targets, the method is general and may be applied to nucleic acid targets as well. Finally, we are likely capturing the free energy components that are less difficult to optimize²³, and quantitative design methods for more difficult components such as electrostatics²⁹ may allow for tighter-binding drugs beyond what we model here. Because our druggability model is physics-based, we can look forward to systematically improving it as our understanding improves or new technologies and approaches are discovered.

METHODS

Binding site structures. Crystal structures were downloaded from the PDB, and inspected for completeness in the binding site. All structures are atomic resolution (<2.5 Å, 1QMF is 2.8 Å) and have a co-crystallized ligand (peptide, small molecule, or substrate). Ligand binding sites were further filled in using MOE SiteFinder alpha-spheres (version 2004.03; Chemical Computing Group). Binding sites were defined by atoms within 5.0 Å of the ligand or alpha spheres and trimmed at the edges to define a contiguous binding site surface of ~300 Å² of surface area.

Calculations. We use computational geometry algorithms previously described^{17,18} to computationally represent binding sites by non-alpha shape Delaunay tetrahedra defined by the binding-site atoms. Surface areas were then computed analytically. This approach allows representation of only the surface that would face a bound ligand. To calculate curvature, we used our previously reported approach of using geometric inversion to find the least-squares fitted (LSF) sphere to the molecular surface, and taking the radius of the sphere as the radius of curvature¹⁷. Our curvature calculation is intuitive but admittedly simple, and we are investigating whether localized curvatures improve the model. Further details on the computational methods can be found in the **Supplementary Methods** online. Conversion of ΔG 's to concentration units is performed using $K_d = \exp\left(-\frac{\Delta G}{RT}\right)$, where $T = 298\text{K}$. Statistical confidence interval for MAP_{POD} values is calculated using free energy (ΔG) values and assuming a two-tailed Student's t distribution. The initial calibration set of eight structures consisted of 1IEP (cAbl kinase), 1KE6 (cyclin dependent kinase 2), 4COX (cyclooxygenase 2), 1HWK (HMG-CoA reductase), 1A4G (neuraminidase), 1OYN (phosphodiesterase 4D), 1MEM (cathepsin K) and 1PTY (phosphotyrosine phosphatase-1B). We note that we are fitting two parameters to eight data points, and as such, the parameters are substantially overdetermined.

HSD and H-PGDS. MAP_{POD} analyses performed using PDB cocrystal structures for HSD (PDB ID: 1EBU) and H-PGDS (PDB ID: 1PD2), and the respective enzyme active sites. An 11,000-compound diversity screening set was screened against *Candida albicans* HSD using an assay that monitors the change in absorbance at 340 nm as a result of NADH to NAD conversion during the homoserine dehydrogenase reaction and using L-aspartate semialdehyde (ASA) as a substrate³¹. The assay was performed with 1 mM ASA ($K_m = 1.05\text{mM}$) and 800 μM NADH ($K_m = 140\text{ }\mu\text{M}$) to identify ASA competitive inhibitors. The same 11,000 compound set was screened against human H-PGDS using an assay that monitors the conversion of prostaglandin H_2 (PGH_2) to PGD_2 based on conversion of unreacted PGH_2 to malondialdehyde (MDA). Recombinant human H-PGDS (10 ng) was incubated in the presence of 20 μM of the substrate, PGH_2 ($K_m = 30\text{ }\mu\text{M}$), in 10 mM potassium phosphate (pH 7.2) and 2.5 mM glutathione ($K_m = 300\text{ }\mu\text{M}$) for 4 min on ice. The incubation was quenched, and unconverted PGH_2 was converted to MDA by incubation with 25 mM FeCl_2 and 50 mM citrate for 30 min. MDA was then reacted with 70 mM thiobarbituric acid to form a fluorescent product ($\text{Ex} = 530\text{ nm}$, $\text{Em} = 550\text{ nm}$), allowing calculation of percent inhibition. For both assays, compounds were tested at 20 μM concentration and were considered 'raw hits' if they produced percent inhibition $\geq 60\%$ in duplicate and had a hit significance of $P \leq 0.05$ based on the distribution of non-hits. Raw hits were followed up with IC_{50} measurements using the same assays.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank colleagues across Pfizer Global R&D for discussions, and Jill Milne and Ralph Lamalot for supporting the HTS experiment. A.C.C. additionally thanks Ken Dill for advice. This work was entirely funded by Pfizer.

AUTHORS CONTRIBUTIONS

A.C.C. developed the method, designed experiments and analyzed data and wrote the manuscript; R.G.C. developed computational geometry algorithms and analyzed data; K.T.S. helped design the HTS experiment, performed screening for HSD, and analyzed results of the screening comparison; P.S. helped design the HTS experiment, developed the H-PGDS assay and performed screening; Q.C. and D.R.C. helped analyze data for Figures 1 and 3a; A.C.S. helped implement computational methods; E.S.H. discussed results and helped design the HTS experiment. All authors reviewed the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/naturebiotechnology/>
Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Brown, D. & Superti-Furga, G. Rediscovering the sweet spot in drug discovery. *Drug Discov. Today* **8**, 1067–1077 (2003).
- Hopkins, A.L. & Groom, C.R. The druggable genome. *Nat. Rev. Drug Disc.* **1**, 727–730 (2002).
- Fauman, E.B., Hopkins, A. & Groom, C.R. Structural Bioinformatics in Drug Discovery. in *Structural Bioinformatics* (eds. Weissig, H. & Bourne, P.) 477–498 (Wiley-Liss, Hoboken, NJ 2003).
- Vassilev, L.T. *et al.* In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
- Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
- Palm, K., Stenberg, P., Luthman, K. & Artursson, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* **14**, 568–571 (1997).
- Veber, D.F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).
- Wermuth, C.G. *The Practice of Medicinal Chemistry*, edn. 2 (Academic Press, London, UK and San Diego, 2003).
- Kuntz, I.D., Chen, K., Sharp, K.A. & Kollman, P.A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. USA* **96**, 9997–10002 (1999).
- Chothia, C. Hydrophobic bonding and accessible surface area in proteins. *Nature* **248**, 338–339 (1974).
- Karplus, P.A. Hydrophobicity regained. *Protein Sci.* **6**, 1302–1307 (1997).
- Sharp, K.A., Nicholls, A., Fine, R.F. & Honig, B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* **252**, 106–109 (1991).
- Cheng, Y.-K. & Rossky, P.J. Surface topography dependence of biomolecular hydrophobic hydration. *Nature* **392**, 696–699 (1998).
- Southall, N.T. & Dill, K.A. The mechanism of hydrophobic solvation depends on solute radius. *J. Phys. Chem. B* **104**, 1326–1331 (2000).
- De Young, L.R. & Dill, K.A. Partitioning of nonpolar solutes into bilayers and amorphous n-Alkanes. *J. Phys. Chem.* **94**, 801–809 (1990).
- Hendsch, Z.S. & Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–226 (1994).
- Coleman, R.G., Burr, M.A., Souvaine, D.L. & Cheng, A.C. An intuitive approach to measuring protein surface curvature. *Proteins* **61**, 1068–1074 (2005).
- Liang, J., Edelsbrunner, H. & Woodward, C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**, 1884–1897 (1998).
- Ettmayer, P., Amidon, G.L., Clement, B. & Testa, B. Lessons learned from marketed and investigational prodrugs. *J. Med. Chem.* **47**, 2393–2404 (2004).
- Kim, C.U. *et al.* Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* **119**, 681–690 (1997).
- Teague, S.J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Disc.* **2**, 527–541 (2003).
- Arkin, M.R. & Wells, J.A. Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nat. Rev. Drug Disc.* **3**, 301–317 (2004).
- Davis, A.M. & Teague, S.J. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Agnew. Chem. Int. Ed.* **38**, 736–749 (1999).
- Nayal, M. & Honig, B. On the nature of cavities on protein surfaces: application to the identification of drug binding sites. *Proteins* **63**, 892–906 (2006).
- Hajduk, P.J., Huth, J.R. & Fesik, S.W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **48**, 2518–2525 (2005).
- Deininger, M., Buchdunger, E. & Druker, B.J. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood* **105**, 2640–2653 (2005).
- Istvan, E.S. & Deisenhofer, J. Structural mechanism for statin inhibition of HMG-CoA reductase. *Science* **292**, 1160–1164 (2001).
- Schachter, M. Chemical, pharmacokinetic and pharmacodynamic properties of statins: an update. *Fundam. Clin. Pharmacol.* **19**, 117–125 (2004).
- Armstrong, K.A., Tidor, B. & Cheng, A.C. Optimal charges in lead progression: a structure-based neuraminidase case study. *J. Med. Chem.* **49**, 2470–2477 (2006).
- Jacques, S.L. *et al.* Characterization of yeast homoserine dehydrogenase, an antifungal target: the invariant histidine 309 is important for enzyme integrity. *Biochim. Biophys. Acta* **1544**, 28–41 (2001).
- Nissink, J.W. *et al.* A new test for validating predictions of protein ligand interaction. *Proteins* **49**, 457–471 (2002).

