2011

A Correlation between Protein Function and Ligand Binding Profiles

Matthew D. Shortridge
University of Nebraska-Lincoln, mds8575@huskers.unl.edu

Michael Bokemper
University of Nebraska-Lincoln

Jennifer C. Copeland
University of Nebraska-Lincoln

Jaime L. Stark
University of Nebraska-Lincoln

Robert Powers
University of Nebraska-Lincoln, rpowers3@unl.edu

Published in final edited form as:
J Proteome Res. 2011 May 6; 10(5): 2538–2545. doi:10.1021/pr200015d.

A Correlation between Protein Function and Ligand Binding Profiles

Matthew D. Shortridge, **Michael Bokemper**, **Jennifer C. Copeland**, **Jaime L. Stark**, and **Robert Powers***
Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304

Abstract

We report that proteins with the same function bind the same set of small molecules from a standardized chemical library. This observation led to a quantifiable and rapidly adaptable method for protein functional analysis using experimentally-derived ligand binding profiles. Ligand binding is measured using a high-throughput NMR ligand affinity screen with a structurally diverse chemical library. The method was demonstrated using a set of 19 proteins with a range of functions. A statistically significant similarity in ligand binding profiles was only observed between the two functionally identical albumins and between the five functionally similar amylases. This new approach is independent of sequence, structure or evolutionary information, and therefore, extends our ability to analyze and functionally annotate novel genes.

Keywords

Protein Function; Ligand Binding; NMR Ligand Affinity Screen; Functional Genomics; Functional Annotation

INTRODUCTION

The recent explosion in sequenced genomes has revealed a vast number of proteins that lack a functional annotation.[1] Many of these unannotated proteins may play an important role in human disease and, correspondingly, are critical for developing new therapeutics. Protein sequence and structure similarity methods are currently the most robust and widely-used tools to annotate a protein of unknown function.[2] Nevertheless, these methods are limited in scope, prone to errors, and based on a small set of experimentally characterized proteins.[3] Only 40 to 60% of sequences suggest a potential functional assignment. Moreover, error rates of < 30% occur even with conservative sequence identities of > 60%. The accuracy of functional annotations decreases substantially in the twilight zone of 20–35% sequence identity.

Recent attempts to extend functional prediction beyond global sequence and structure similarity have led to the development of active-site similarity search methods.[4–7] These methods try to identify protein surface structures that interact with biologically important ligands since active-sites that share a similarity in sequence, structure and ligand binding are

*To whom correspondence should be addressed: Department of Chemistry, 722 Hamilton Hall, University of Nebraska – Lincoln, Lincoln, NE 68588-0304, Tel: (402) 472-3039; Fax: (402) 472-9402, rpowers3@unl.edu.

SUPPORTING INFORMATION AVAILABLE

Supporting Information Available: three supplementary tables that include the complete summary of the NMR ligand affinity screen for all 19 proteins, all pairwise functional similarity scores and all pairwise ligand binding profile similarity scores is available free of charge via the Internet at http://pubs.acs.org.

predicted to be functionally related. This is based on the fundamental principal that a protein's active-site has been optimized by nature to interact with a unique and specific set of targets, where this information can be leveraged to understand function. Consequently, protein surfaces have been shown to be exquisitely selective and to only bind ligands at very specific functionally relevant locations.[8–11] This understanding is also essential to drug discovery, where extensive resources are allocated by the pharmaceutical industry to identify high-affinity and selective compounds that target a specific therapeutically relevant protein.[12, 13] The use of ligands as functional probes is the basis of our FAST-NMR methodology[4, 14] that has been sucessfuly applied to explore the function of *Staphylococcus aureus* protein SAV1430,[4] *Pseudomonas aeruginosa* protein PA1324,[15] *Pyrococcus horikoshii* OT3 protein PH1320,[14] human protein Q13206,[14] *Bacillus subtilis* protein YndB,[16] and *Salmonella typhimurium* PrgI protein.[17] Similar successes have been reported using ligand binding to infer function in virtual screens.[18, 19] While promising, current active-site similarity techniques still rely on high-resolution protein structures to identify and measure functional similarity.[20] The availability of structures for the entire proteome remains a significant bottleneck for the high-throughput functional annotation of hypothetical proteins.

We report herein a new method to infer protein function that is independent of sequence and structural information. Our method uses a similarity in ligand binding profiles to annotate a protein of unknown function. This is similar in concept to the mapping of pharmacological space or the use of structure-activity relationships (SAR) for target selection and chemical lead identification in drug discovery.[21–24] A ligand binding profile is defined as a set of ligands that bind a protein from a high-throughput ligand affinity screen. Ligand binding is monitored using our 1D ^{1}H NMR line-broadening screen.[25] In essence, the chemical and structural diversity of a compound library provides an experimental means of mapping the physiochemical properties of a protein's active-site based on the compounds that do or do not bind the protein. Functional annotation is inferred by clustering unknown proteins with previously annotated proteins that share similar ligand binding profiles from the same chemical library. A modification of the E-value routinely used in sequence homology is used to quantify ligand binding profile similarities. The methodology is demonstrated using 19 proteins with a range of function defined by Gene Ontology (GO) terms.[26]

EXPERIMENTAL

Materials

The human serum albumin (HSA) (essentially fatty acid free, ≥ 96 % pure), bovine serum albumin (BSA) (minimum 98% agarose gel electrophoresis, lyophilized), α-amylase from *Bacillus lincheniformis* (Bli) (500–1,500 units/mg protein, 93–100% (SDS page)), α-amylase from *Aspergillus oryzae* (Aor) (powder, ~30 units/mg), α-amylase from *Bacillus amyloliquefaciens* (Bam) (liquid, ≥250 units/g protein), β-amylase from barley (Hvu) (type II-B 20–80 units/mg protein), and β-amylase from sweet potato (Iba) (Type I-B, ammonium sulfate suspension, ≥750 units/mg protein) protein samples were all purchased from Sigma (St. Louis, MO). The *S. typhimurium* PrgI protein samples and assigned ^{1}H-^{15}N HSQC spectrum were generously provided by Dr. Roberto DeGuzman (University of Kansas). *Staphylococcus aureus* primase C-Terminal domain (CTD) protein sample was purchased from Nature Technologies Corporation (Lincoln, NE). *H. sapiens* diacylglycerol kinase alpha (DGKA), *P. aeruginosa* unannotated protein PA1324, *S. aureus* unannotated protein SAV1430, *S. typhimurium* unannotated protein STM1790, *H. sapiens* ubiquitin-fold modifier-conjugating enzyme 1 (UFC1), *E. coli* unannotated protein YjbR, *E. coli* unannotated protein YkfF, *B. subtilis* unannotated protein YkvR and *E. coli* unannotated protein YtfP protein samples were provided by Dr. Gaetano Montelione, Director of the Northeast Structural Genomics Consortium (NESG, www.nesg.org). The *S. aureus* nuclease

was over-expressed in house from a cell stock of *E. coli* Bl21 DE3 codon+ (Stratagene) containing the pET28(a)+plasmid with the *dnuc* gene provided by Dr. Greg Somerville (University of Nebraska-Lincoln) grown in LB broth and purified using a Talon cobalt affinity resin (Clontech). The deuterium oxide (99.9 atom% D) and the dimethyl sulfoxide-d_6 (99.9% D) were purchased from Aldrich (Milwaukee, WI) The 3-(trimethylsilyl)propionic acid-2,2,3,3-d_4 (TMSP-d_4) was purchased from Cambridge Isotope (Andover, MA). The Bis-Tris-d_{19} (98 atom% D CP) was purchased from Isotec (Milwaukee, WI). The compound library was previously complied as described elsewhere.[27]

NMR Data Collection and Sample Preparation

All NMR data was collected on a Bruker 500 MHz Avance spectrometer (Billerica, MA) equipped with a triple resonance, Z-axis gradient cryoprobe and using a Bruker BACS-120 sample changer and IconNMR software for automated data collection. The screening data for this study was compiled over a 5 year time span in which two different 1D ^{1}H solvent suppression pulse sequences were used for the measurement of ligand 1D ^{1}H NMR line broadening. Data for the HSA, BSA, *S. aureus* primase CTD, PrgI, PA1324, and SAV1430 were collected as previously described.[4, 15, 17, 25] Data for DGKA, STM1790, UFC1, YjbR, YkfF, YkvR and YtfP, the 5 amylases and *S. aureus* nuclease proteins was collected at 298 K using 64 transients with a spectrum width of 6009 Hz with 8 K data points and a 1.0 sec relaxation delay using the excitation sculpting[28] method for solvent suppression of the residual H_2O resonance signal. The samples for the HSA, BSA, *S. aureus* primase CTD, PrgI, PA1324, and SAV1430 NMR screens were prepared as previously described.[4, 15, 17] *S. aureus* nuclease, DGKA, STM1790, UFC1, YjbR, YkfF, YkvR, YtfP, and the 5 amylases were screened at 5 μM protein concentration and 100 μM ligand concentration in a screening buffer of 2% DMSO-d6, 20 mM Bis-Tris pH 7.0, 11.1 mM TMSP-d4 in “100%” D_2O.

Chemical Library

All NMR ligand affinity assays were completed by screening each protein individually with a library of 437 biologically active compounds (http://bionmr-c1.unl.edu/ligands).[27] The library contains amino-acids, carbohydrates, co-factors, fatty-acids, hormones, inhibitors, known drugs, metabolites, neurotransmitters, nucleotides, and substrates. The compound library is divided into 116 mixtures with 3–4 ligands per mixture and is described in detail elsewhere. In order to assess the structural diversity of the library, 1300 molecular descriptors were calculated for each compound using the online software eDragon (VCClabs, http://www.vcclab.org/lab/edragon/).[29] MM2 minimized 3D MOL2 files were generated using ChemBio 3D Ultra 12.0 (CambridgeSoft, Cambridge, MA), converted to SMILES using OpenBabel (http://openbabel.org) and then uploaded to the eDragon web site. The molecular descriptors calculated for each structure were incorporated into a single Excel spreadsheet and imported into SIMCA (UMETRICS, Kinnelon, NJ). Each molecular descriptor was treated as separate bin or data point for each structure. A 3D PCA scores plot was generated using the calculated molecular descriptors for the structures in the library.

False positive and false negative rates were simulated to determine if the screening library of 437 compounds is of sufficient size to make meaningful comparisons between proteins of unknown function. An in-house program was written that randomly generates a ligand binding profile using a Gaussian distribution about two means: (i) average hit rate of 32 ± 44 bound ligands, or (ii) a lower hit rate of 16 ± 6. Either 1×10^6 random pairs of ligand binding profiles were generated or a single randomly generated ligand binding profile was compared against a random set of 1×10^6 ligand binding profiles. The simulations were done in triplicate and the library sizes used in the simulations corresponded to 437, 1,000, 2,000, 5,000, and 10,000 compounds. An E-value of $\leq 1\times10^{-9}$ was used to define a similar ligand

binding profile. A histogram of the Log(E-values) were plotted and fitted using EasyFit V5.4 (MathWave Technologies).

To estimate a false negative rate, an error was introduced to randomly generated pairs of identical ligand binding profiles. Each ligand binding profile has false binders added or true binders removed at a percentage of the rate that a true binder was added to the original ligand binding profile (based on the original number of predicted binders (m and n) chosen from the Gaussian distribution):

$$m_e = m_o \pm e m_o \text{ and } n_e = n_o \pm e n_o \quad (1)$$

where e is the error rate (10%-50%), m_e and n_e ($m_e \neq n_e$) are the new number of bound ligands after the error rate is applied, and m_o and n_o ($m_o = n_o$) are the original number of bound ligands predicted from the Gaussian distribution.

Binding Assay

Ligand binding was manually identified from a decrease in the free ligand 1D ^{1}H NMR signal upon the addition of protein. This decrease is determined by visually comparing ligand peak intensities to the TMSP-d_4 methyl resonance (0.00 ppm) from the 100 μM TMSP-d_4 internal standard. Any ligand with a visually observable decrease in peak height from the addition of a protein is considered to be a binder. A detailed analysis of the relationship between K_D and NMR line-broadening has been previously discussed in detail.[30] From this analysis, a conservative estimate of our limit of detection can be made, which corresponds to ligands with a K_D of > 100–300 μM. Of course, this limit is dependent on the molecular-weight of the protein, where sensitivity increases with MW. Thus, our ligand binding assay will be dominated by biologically relevant protein-ligand interactions, where non-specific or irrelevant interactions start to dominate as the K_D increases beyond 300 μM.[31] Conversely, tight-binders ($K_D \leq$ nM) that are governed by slow-off rates may simply result in a decrease in peak intensity proportional to the limiting protein concentration. A 5% change in peak intensity may be difficult to decipher and correspond to a false positive. Nevertheless, encountering tight binders in ligand binding assay is generally a rare event. Binders from our chemical library are typically structural homologs to the natural ligand. Also, these tight binders would be expected to be uniformly missed for functionally similar proteins. The methods for data processing and identifying binding ligands have been previously discussed in detail.[25, 27, 30] Overall, for our library of 437 compounds, the 1D ^{1}H NMR line-broadening screen requires approximately a day to complete both the data acquisition and the data analysis.

Ligand Binding Profiles

A similarity in ligand binding profiles was measured between each pair of proteins using equation 1. Overlapping binding ligands (S) for every protein in a pairwise manner were identified by comparing a list of all binding ligands and counting the number of overlapping ligands. Each pairwise E-value was calculated using a library size of 437 compounds (p = 1/437 = 0.00229). An Excel spreadsheet program was written to match overlapping ligands and measure E-values.

Functional Similarity Measurement

The Uniprot accession number was obtained for each protein in the study. The list of Uniprot accession numbers was uploaded to the semantic similarity tool FunSimMat (http://funsimmat.bioinf.mpi-inf.mpg.de/). All reported functional similarities are expressed as a *funsim* score measured as previously described.[32]

RESULTS AND DISCUSSION

Structural Diversity of the Screening Library

Our chemical library for NMR ligand affinity screening was designed to maximize functional diversity.[27] In addition to practical considerations such as solubility, stability and cost, compounds were added to our library based on a known biological activity involving a distinct protein or protein class. Compounds correspond to known drugs, inhibitors, substrates or cofactors. Not surprisingly, the compounds are also consistent with typical "drug-like" characteristics and with fragment libraries.[33, 34] These characteristics include good aqueous solubility, low molecular-weights, and low number of rings, heteroatoms, and hydrogen-bond donors and acceptors. Diversity in biological activity was also anticipated to result in a correlated diversity in chemical structure. To validate the structural diversity of our functional chemical library, ~1,300 different molecular descriptors were calculated for each compound.[29] A principal component analysis (PCA) of the set of molecular descriptors indicates a uniform coverage of structural space. A 3D PCA scores plot is shown in Figure 1A. The structures are distributed throughout the structural space defined by the molecular descriptors. Conversely, if there was an overabundance of any structural class, distinct clustering patterns would be apparent in the 3D PCA scores plot. Clearly, our chemical library is an acceptable set of molecular probes to evaluate a diversity of protein function.

Calculation of Ligand Binding Profile Similarities

Measuring a significant similarity between two ligand binding profiles requires the development or adaptation of a robust scoring function. Current similarity scoring methods used for sequence analysis, such as the E-value developed by Karlin and Altschul,[35] are also well-suited for measuring a similarity between ligand binding profiles.

$$E = Kmne^{-\lambda S} \tag{2}$$

Here, the E-value is only dependent on the total number of compounds that bind each protein (m and n) and the total number of compounds that bind both proteins (S). Additionally, the probability of finding a significant similarity is proportional to the probability search space (K) and scoring function (λ).

$$K = \frac{(q - p')^2}{q} \text{ and } \lambda = \ln \frac{q}{p'} \tag{3}$$

Unlike sequence similarity, a similarity between ligand binding can be thought of as a binary system (binding vs. non-binding) therefore the probabilities p′ and q simply becomes the probability of finding a hit within a library:

$$p' = \frac{1}{\text{library size}} \tag{4}$$

and the probability of finding a ligand that binds both proteins:

$$q = \frac{S}{m * n} \tag{5}$$

The standard E-value also provides a robust measure of the probability that the ligand binding similarity is not due to chance using the standard P-value.

$$P=1-e^{-E} \tag{6}$$

As expected, the ligand binding profile E-value rapidly becomes insignificant (P > 0.0001) as the probability of finding a ligand that binds both proteins (q) decreases. Binding profiles that have a P < 0.0001 are significant at the 99.99% confidence interval ($E=10^{-5}$). Thus, our method is only dependent on comparing the total number of binding events (m or n) and the set of overlapping binding ligands (S) between two proteins.

Sufficient Size of a Screening Library

Obtaining a balance between library depth and breadth is very challenging and has been a focus of compound library design for over a decade – without a clear consensus conclusion.[36] Clearly, the size of the library would be expected to impact the number of observed binders (m and n) and the corresponding similarity in ligand binding profiles (S and E-value). Fundamentally, determining the optimal size of the chemical library is an open-ended, and at some level, a very difficult question to adequately answer. It is always plausible for a protein to be screened that results in a complete absence of binders regardless of the size or composition of the chemical library. If the protein is a true unknown, how is it possible to ascertain *a priori* that the library composition is adequate? The only recourse is to explore the probability of identifying binders within a given set of reasonable assumptions and given experimental hit rates.

On average, 32 ± 44 ligands were observed to bind a protein target in our NMR ligand affinity screen. Our simulations indicate that even with a modest library size of 437 compounds, the probability of randomly finding two similar (E-value ≤ 1×10^{-9}) ligand binding profiles was shown to be effectively zero. This is not too surprising considering that in theory there are 2^{437} (3.5×10^{131}) different binding profiles, where the product (1.3×10^{263}) leads to an effectively miniscule probability of finding two similar ligand binding profiles. Of course, only a small subset of these potential ligand binding profiles are possible given 32 ± 44 bound ligands, but this still represents a very large number of dissimilar pairs of ligand binding profiles. A randomly selected ligand binding profile using a Gaussian distribution of bound ligands with a smaller mean (larger potential false positive rate) of 16 ± 6 was compared against a random set of 1×10^{6} ligand binding profiles using the same Gaussian distribution to select binders. A histogram of the Log(E-values) is shown in Figure 1B and best fitted with the Weibull Distribution (Extreme Value Type III Distribution), which indicates the calculated E-values are significant.[37] Consequently, the comparison did not yield any significant similarities and the most common occurrence was an overlap (S) of zero (S ranged from 0 to 7).

While the false positive rate is effectively zero, a false negative rate was measurable and, as expected, decreased for increasing library size. Again, a total of 1×10^{6} pairs of *identical* ligand binding profiles (m=n=S) was randomly generated using a Gaussian distribution with a mean of 16 ± 6 bound ligands (m and n). An error rate ranging from 10%–50% was introduced into each ligand binding profile, independently changing the two identical ligand binding profiles. The simulations were repeated for library sizes that ranged from 437 to 10,000 compounds. The percentage of false negatives (E-value of > 1×10^{-9}) found in each simulation are plotted as a function of library size in Figure 1C. The false negative rate increases proportional to the error rate and decreases proportional to the library size. For our library of 437 compounds, the percentage of false negatives is ~9% with a 50% error rate

(see eqn. 1). Conversely, only a ~2% false negative rate is observed for a library of 2,000 compounds at the maximum error rate of 50%. The false negative rate is below 1% for a library of 10,000 compounds. Correspondingly, ligand-binding profile similarities are relatively tolerant to erroneous binders. This is consistent with the lack of any false negatives in the 19 screens reported herein. Thus, the simulations indicate that even a modest library of 437 compounds provides a relatively robust and reliable measure of functional similarity, but a slight increase in the library size may improve the methods accuracy. Of course, increasing the library size also increases assay time, but a library of 1500–2000 compounds is still practical since the assay time is only estimated to increase to ~1.5–2 days.

The library size also defines the minimal number of binders (m, n) and overlapping binders (S) required for obtaining a significant E-value of 1×10^{-9}. For a modest library of 437 compounds, the minimal number of binders and overlapping binders is 5 compounds. The number drops to 4 compounds for a library size of 1000–2000 compounds and to 2 compounds for a library size of 5000–10000. Considering the average number of binders is 32 ± 44, these are effectively inconsequential improvements for a substantial increase in screening time. Alternatively, false negatives in the binding assay (missed tight binders) may be potentially detrimental to proteins that bind a very limited number of ligands (< 5). In principal, a single false negative may be the difference between a significant or insignificant E-value. Of course, the number of binders is expected to scale with the library size, assuming a relatively constant hit rate.[38] Correspondingly, increasing the library size to 1500–2000 compounds is expected to make proteins that bind only four or less ligands a relatively rare event.

Correlating Protein Function with Ligand Binding Profiles

To experimentally support the ligand binding profile hypothesis, 19 proteins were screened by NMR using our chemical library of biologically active compounds.[27] Binding events were identified as previously described by measuring a decrease in ligand ^{1}H NMR peak intensities in the presence of a protein (Figure 2a).[4, 25] Thus, the ligand binding profile is simply a binary list that indicates which compounds out of the library of 437 compounds were shown to bind the protein. The complete summary of results from the NMR ligand affinity screen for the 19 proteins can be found in Table 1S.

For the 19 proteins screened in the NMR ligand affinity assay, 13 proteins have a previously annotated function based on GO terms and 6 proteins have an unknown function. The 19 proteins were chosen to contain two sets of functionally similar proteins mixed with a third set of functionally diverse proteins. The two sets of functionally related proteins are 2 serum albumins and 5 amylases. The serum albumins and amylases were chosen because the proteins have a function related to ligand binding and were readily available from commercial sources. The additional 12 proteins are from NESG or other on-going functional annotation projects involving our FAST-NMR methodology.[4, 14] The primary intent of these additional proteins is to provide a “functional background” to test the ability of the ligand binding profile to distinguish the serum albumins and amylases from each other and from the remaining proteins. Will the addition of the 12 functionally diverse proteins cause erroneous similarities to the albumins or amylases that is not correlated with function?

A FunSimMat functional similarity score was calculated for each pair of proteins within the set of 19 proteins.[32] FunSimMat uses GO terms to generate a semantic similarity score that ranges from 0 for no functional similarity to 1 for identical functions. An average FunSimMat similarity score of 0.98 and 0.67 ± 0.04 was calculated between the albumins and amylases, respectively. The remaining 12 proteins exhibited no functional relationship to any other protein in the screening set, yielding an average FunSimMat similarity score of

0.1 ± 0.1. The complete list of FunSimMat similarity scores can be found in Table 2S. A weak functional similarity was observed between the two albumins and the human protein ubiquitin-fold modifier-conjugating enzyme 1 (UFC1, Uniprot: Q9Y3C8). However, this similarity is limited to one overlapping and generic "protein binding" GO number (GO: 0005515).

An all-vs-all pairwise comparison of the 19 ligand binding profiles gave a total of 171 ligand binding profile comparisons with only 11 comparisons giving a significant similarity score ($P < 0.0001$). The comparisons with the highest similarity scores corresponded to the set of albumins (E-value 1×10^{-58}) and the set of amylases (median E-value 3×10^{-13}). Conversely, the median E-value for the remaining ligand binding comparisons was 0.1. For comparison, a median E-value of 3×10^{-26} was obtained when all the ligand binding profiles are compared to themselves. These results clearly indicate that ligand binding profile similarities are strongly correlated with functional similarity. All the protein pairs with a significant ligand binding similarity score along with the corresponding FunSimMat functional similarity score are listed in Table 1. The complete list of ligand binding similarity scores can be found in Table 3S. It is also important to note that the absolute magnitude of a ligand binding profile E-value is directly dependent on the total number of ligands shown to bind a protein. This is equivalent to sequence homology where the E-value scales by the length of the protein sequences.

The overall similarity in the ligand binding profiles is also easily visualized in a heat map (Figure 2b). A ligand identified by NMR to bind a protein is simply indicated by a red line in the heat map. The ligands are sorted first by their ability to bind human serum albumin (HSA) and then by their binding to *A. oryzae* α-amylases (Aor). An expansion of the heat map focused on the amylases and sorted by ligand binding to Aor is also shown in Figure 3. The heat map clearly shows overlapping clusters of ligands between the albumins and amylases. The remainder of proteins exhibits no similarity in the ligand binding profile based on the obvious random scatter in the heat map.

There was also a minimal similarity in ligand binding between *S. aureus* nuclease and the α-amylases from *A. oryzae* and *B. amyloliquefaciens* (median E-value 4×10^{-5}). It is plausible that this minimal similarity is simply due to a serendipitous overlap in non-specific ligand binding between the three proteins. However, the similarity in the ligand binding profiles was limited to the nucleosides in the library. Additionally, the remaining 3 amylases did not bind these ligands or exhibit a significant ligand binding similarity to nuclease. The observed ligand binding similarity between the nuclease and two of the α-amylases is potentially due to trace amounts of a nuclease that may be present in the *A. oryzae* and *B. amyloliquefaciens* α-amylases samples. This is a likely occurrence since the samples were purchased as crude mixtures, where size-exclusion chromatography only yielded a modest improvement in purity. This illustrates an important consideration in the general application of ligand binding profiles. False positives in the ligand affinity assay due to impurities, non-specific binding, or experimental concerns (precipitation, aggregation, etc.) may lead to an inaccurate functional assignment. Proper care in the execution and analysis of ligand binding profiles should minimize these concerns.

As shown in table 1, HSA and BSA had a large number of binding ligands (178 and 171, respectively) compared to the overall size of the library. The relative hit rate for these two proteins was 40.7% and 39.1% respectively. With a large hit rate, false similarities may arise if a second protein serendipitously binds to a small subset of compounds that were shown to bind HSA or BSA. However, the ligand binding similarity score (eqn. 2) effectively eliminates this concern by scaling the score based on both the total number of compounds found to bind each protein and by the number of overlapping binding ligands. As an

example, the *S. typhimurium* type III secretion system protein PrgI bound to a total of five compounds, where each compound was also shown to bind HSA and BSA. The corresponding E-values for the ligand binding profile comparisons between PrgI and HSA (7×10^{-2}) and BSA (6×10^{-2}) were not significant at a P=0.0001.

Ligand binding profiles are independent of sequence and structural information and thus provide an experimentally based approach to predict protein function in a relatively robust and high-throughput fashion. The results reported herein demonstrate a clear correlation between ligand binding similarity scores and FunSimMat functional similarity scores. Specifically, only the set of albumins and amylases gave significant ligand binding similarity scores. Unfortunately, the ligand binding profiles were unable to differentiate between the two α and β amylase families. A further refinement of the functional annotation would require a second screening step using a focused library to differentiate these functional classes. In the case of the amylases, this would involve screening the proteins with a carbohydrate library, where a subset of the compounds would selectively bind to the α- or β-amylase proteins. Alternatively, a larger chemical library with an increase in the number of compounds per representative class, such as additional carbohydrates, would be expected to enhance the functional resolution of the technique.

While our methodology has been shown to be effective with the proteins examined, limitations may be encountered with other classes of proteins. An NMR ligand affinity screen using intrinsically disordered proteins would be unproductive unless ligand binding induced a folded state or a binding partner that stabilized a folded state was present. Of course, the presence of a binding partner would complicate the data analysis; does the ligand bind the complex or binding partner instead of the targeted protein? Membrane proteins would be equally challenging, requiring methods to prepare adequate quantities of the protein for the NMR screen while requiring lipid bicelles, micelles or detergents to stabilize the protein. A similar data analysis problem would arise. Do the ligands interact with the lipid bicelles, micelles or detergents instead of or in addition to the protein target? Furthermore, does the NMR sample preparation procedure affect the solubility or aggregation state of compounds in the library? Finally, proteins that bind a very limited number of compounds from our library (< 5) would result in a ligand binding profile that would only yield insignificant E-values ($> 1\times10^{-9}$). Despite these potential limitations and challenges, ligand binding profiles are expected to be broadly applicable to the majority of the proteome.

CONCLUSION

The success of whole-genome sequencing has generated an enormous dataset of functionally uncharacterized proteins. Sequence and structure homology are routinely used to leverage functional annotations, but > 30% of the proteome lack a sequence or structure similarity to proteins of known function. Alternatively, detailed experimental analysis may require upwards of a decade of effort to characterize a single protein. Instead, we describe the use of high-throughput (HTS) NMR ligand affinity screens to infer a biological function through a similarity in ligand binding profiles. A diverse chemical library is used to map the physiochemical properties of a protein's active-site, where the identity of the ligands that bind a protein provides information about the biological activity of the protein. A modification to the E-value developed by Karlin and Altschul allows for a similarity between ligand-binding profiles to be measured, where an E-value $\leq 1\times10^{-9}$ suggests functional homologs. We demonstrated that the preponderance of binding ligands identified from 19 NMR ligand affinity screens were uniquely associated with each functional class and were shown to correlate with the protein's function based on GO terms (Figure 2b and Table 1).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Dr. Roberto De Guzman from the University of Kansas for supplying the PrgI sample, and Dr. Gaetano Montelione, Director of the Northeast Structural Genomics Center (NESG, www.nesg.org) for supplying the DGKA, PA1324, SAV1430, STM1790, UFC1, YjbR, YkfF, YkvR and YtfP protein samples for the NMR ligand affinity screen. We would like to thank Dr. Greg Somerville from the University of Nebraska-Lincoln for providing the *S. aureus* nuclease expression vector and Drs. Byron Taylor and Rathnam Chaguturu from the University of Kansas for providing numerous high-throughput screening data sets. This work was supported by the National Institute of Allergy and Infectious Diseases Nebraska [R21AI081154], and by the Tobacco Settlement Biomedical Research Development Fund. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Allergy and Infectious Diseases. Research was performed in facilities renovated with support from the National Institutes of Health [RR015468-01].

References

1. Janitz M. Assigning functions to genes--the main challenge of the post-genomics era. Rev Physiol Biochem Pharmacol. 2007; 159:115–29. [PubMed: 17846923]
2. Rentzsch R, Orengo CA. Protein function prediction--the power of multiplicity. Trends Biotechnol. 2009; 27(4):210–9. [PubMed: 19251332]
3. Rost B, Liu J, Nair R, Wrzeszczynski KO, Ofran Y. Automatic prediction of protein function. Cell Mol Life Sci. 2003; 60(12):2637–50. [PubMed: 14685688]
4. Mercier KA, Baran M, Ramanathan V, Revesz P, Xiao R, Montelione GT, Powers R. FAST-NMR: functional annotation screening technology using NMR spectroscopy. J Am Chem Soc. 2006; 128(47):15292–9. [PubMed: 17117882]
5. Kinnings SL, Jackson RM. Binding site similarity analysis for the functional classification of the protein kinase family. J Chem Inf Model. 2009; 49(2):318–29. [PubMed: 19434833]
6. Ferre F, Ausiello G, Zanzoni A, Helmer-Citterich M. Functional annotation by identification of local surface similarities: A novel tool for structural genomics. BMC Bioinf. 2005; 6 Epub.
7. Skolnick J, Brylinski M. FINDSITE: a combined evolution/structure-based approach to protein function prediction. Briefings Bioinf. 2009; 10(4):378–391.
8. Hajduk PJ, Huth JR, Fesik SW. Druggability Indices for Protein Targets Derived from NMR-Based Screening Data. J Med Chem. 2005; 48(7):2518–2525. [PubMed: 15801841]
9. English AC, Groom CR, Hubbard RE. Experimental and computational mapping of the binding surface of a crystalline protein. Protein Eng. 2001; 14(1):47–59. [PubMed: 11287678]
10. Allen KN, Bellamacina CR, Ding X, Jeffery CJ, Mattos C, Petsko GA, Ringe D. An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins. J Phys Chem. 1996; 100(7): 2605–11.
11. Liepinsh E, Otting G. Organic solvents identify specific ligand-binding sites on protein surfaces. Nat Biotechnol. 1997; 15(3):264–268. [PubMed: 9062927]
12. Kubinyi H. Structure-based design of enzyme inhibitors and receptor ligands. Curr Opin Drug Discovery Dev. 1998; 1(1):4–15.
13. Gubernator K, Boehm HJ. Examples of active areas of structure based-design. Methods Princ Med Chem. 1998; 6:15.
14. Powers R, Copeland J, Mercier K. Application of FAST-NMR in Drug Discovery. Drug Discov Today. 2008; 13(3–4):172–179. [PubMed: 18275915]
15. Mercier KA, Cort JR, Kennedy MA, Lockert EE, Ni S, Shortridge MD, Powers R. Structure and function of Pseudomonas aeruginosa protein PA1324 (21–170). Protein Sci. 2009; 18(3):606–18. [PubMed: 19241370]
16. Stark JL, Mercier KA, Mueller GA, Acton TB, Xiao R, Montelione GT, Powers R. Solution structure and function of YndB, an AHSA1 protein from Bacillus subtilis. Proteins Struct Funct Bioinf. 2010; 78(16):3328–3340.

17. Shortridge MD, Powers R. Structural and functional similarity between the bacterial type III secretion system needle protein PrgI and the eukaryotic apoptosis Bcl-2 proteins. PLoS One. 2009; 4(10):e7442. [PubMed: 19823588]

18. Keiser MJ, Roth BL, Armburster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. Nat Biotechnol. 2007; 25(2):197–206. [PubMed: 17287757]

19. Xie L, Li J, Xie L, Bourne PE. Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. PLoS Comput Biol. 2009; 5(5):e1000387. [PubMed: 19436720]

20. Dessailly BH, Lensink MF, Orengo CA, Wodak SJ. LigASite--a database of biologically relevant binding sites in proteins with known apo-structures. Nucleic Acids Res. 2008; (Database issue):D667–73. [PubMed: 17933762]

21. Paolini GV, Shapland RHB, van HWP, Mason JS, Hopkins AL. Global mapping of pharmacological space. Nat Biotechnol. 2006; 24(7):805–815. [PubMed: 16841068]

22. Vieth M, Sutherland JJ, Robertson DH, Campbell RM. Kinomics: Characterizing the therapeutically validated kinase space. Drug Discov Today. 2005; 10(12):839–846. [PubMed: 15970266]

23. Bamborough P, Drewry D, Harper G, Smith GK, Schneider K. Assessment of Chemical Coverage of Kinome Space and Its Implications for Kinase Drug Discovery. J Med Chem. 2008; 51(24): 7898–7914. [PubMed: 19035792]

24. Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJ Jr, Kohn KW, Fojo T, Bates SE, Rubinstein LV, Anderson NL, Buolamwini JK, van OWW, Monks AP, Scudiero DA, Sausville EA, Zaharevitz DW, Bunow B, Viswanadhan VN, Johnson GS, Wittes RE, Paull KD. An information-intensive approach to the molecular pharmacology of cancer. Science. 1997; 275(5298):343–349. [PubMed: 8994024]

25. Mercier KA, Shortridge MD, Powers R. A multi-step NMR screen for the identification and evaluation of chemical leads for drug discovery. Comb Chem High Throughput Screen. 2009; 12(3):285–95. [PubMed: 19275534]

26. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25(1):25–9. [PubMed: 10802651]

27. Mercier KA, Germer K, Powers R. Design and characterization of a functional library for NMR screening against novel protein targets. Comb Chem High Throughput Screen. 2006; 9(7):515–34. [PubMed: 16925512]

28. Hwang T-L, Shaka AJ. Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients. J Magn Reson, Ser A. 1995; 112(2):275–9.

29. Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin VA, Radchenko EV, Zefirov NS, Makarenko AS, Tanchuk VY, Prokopenko VV. Virtual computational chemistry laboratory--design and description. J Comput Aided Mol Des. 2005; 19(6):453–63. [PubMed: 16231203]

30. Shortridge MD, Hage DS, Harbison GS, Powers R. Estimating protein-ligand binding affinity using high-throughput screening by NMR. J Comb Chem. 2008; 10(6):948–58. [PubMed: 18831571]

31. Breuker K. The study of protein-ligand interactions by mass spectrometry-a personal view. Int J Mass Spectrom. 2004; 239(1):33–41.

32. Schlicker A, Albrecht M. FunSimMat: a comprehensive functional similarity database. Nucleic Acids Res. 2008; (Database issue):D434–9. [PubMed: 17932054]

33. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Delivery Rev. 1997; 23(1–3):3–25.

34. Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discovery. 2007; 6(3):211–219.

35. Karlin S, Altschul SF. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc Natl Acad Sci U S A. 1990; 87(6):2264–8. [PubMed: 2315319]

36. Dolle RE. Historical overview of chemical library design. Methods Mol Biol. 2010; 685(1):3–25. [PubMed: 20981516]

37. Baldi P, Nasr R. When is Chemical Similarity Significant? The Statistical Distribution of Chemical Similarity Scores and Its Extreme Values. J Chem Inf Model. 2010; 50(7):1205–1222. [PubMed: 20540577]

38. Chen IJ, Hubbard RE. Lessons for fragment library design: analysis of output from multiple screening campaigns. J Comput-Aided Mol Des. 2009; 23(8):603–620.

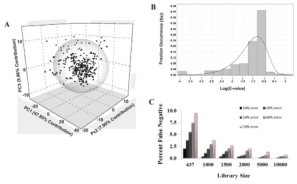

Figure 1.

(A) Three dimensional PCA scores plot where each point represents one compound from the functional chemical library. The placement of each point in the PCA scores plot is indicative of the unique structural identity for each compound. The contribution of each principal component is labeled on the axis. The sphere represents the 95% confidence limit. **(B)** A histogram distribution of E-values calculated from a simulation of ligand binding profiles. A random ligand-binding profile was compared against a random set of 1×10^6 ligand binding profile using a library of 437 compounds. The solid line corresponds to the best fit curve from the Weibull Distribution (Extreme Value Type III Distribution) model. **(C)** A plot of the percentage of false negatives as a function of error rate (10% to 50%) and library size (437 to 10,000) from a simulation of ligand-binding profiles.

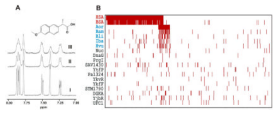

Figure 2.

(A) Ligand binding is identified by a decrease in ligand peak intensity upon addition of a target protein. The 1D ^{1}H NMR spectrum of the non-steroidal anti-inflammatory drug naproxen **(I)** is shown to broaden in the presence of *H. sapiens* serum albumin (HSA) **(II)** and *B. Taurus* serum albumin (BSA) **(III)** indicating a positive binding event. The NMR line broadening experiments used 100 μM ligand and 5 μM protein as described in the methods section. **(B)** Heat map summarizing the NMR ligand affinity screens for 19 proteins: where the albumins are colored red, the amylases cyan and the remainder of the proteins gray. A binding ligand is indicated by a red line. The 437 ligands were sorted to maximize the clustering of binding ligands for the albumins and amylases.

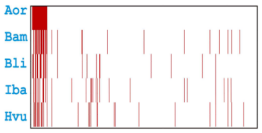

Figure 3.
Expanded view of the heat map shown in Figure 2 highlighting the similarity in ligand binding profiles for the amylases. The 437 ligands were sorted to maximize the clustering of binding ligands for the amylases.

Supplementary Tables for "A Correlation between Protein Function and Ligand Binding Profiles" Matthew D. Shortridge, Michael Bokemper, Jennifer C. Copeland, Jamie Stark and Robert Powers*

Supplementary Table 1S: Summary of NMR Ligand Affinity Screening Results[a]

molID	commonName	plate	well	priMix	HSA	BSA
2	Adenosine 5-triphosphate disodium salt (ATP)	1	A1	1001	1-Methylimidazole	1-Methylimidazole
3	Cytidine 5'-triphosphate disodium salt (CTP)	1	B1	1002	5-Fluoro-5'-deoxyuridine	5-Fluoro-5'-deoxyuridine
4	Guanosine 5'-triphosphate sodium salt hydrate (GTP)	1	C1	1003	(-)-Perillic acid	(-)-Perillic acid
5	Thymidine 5'-triphosphate sodium salt (TTP)	1	D1	1004	(±)-2,3-Dichloro-a-methylbenzylamine hydrochloride	[Hydroxy(tosyloxy)iodo]benzene
6	Uridine 5'-triphosphate trisodium salt dihydrate (UTP)	1	E1	1005	(±)-Carnitine chloride	1-(4-Chlorobenzyl)-5-methoxy-2-methylindole-3-acetic acid
7	Adenosine 3',5'-cyclic monophosphate (cyclic AMP)	1	F1	1006	(±)-Propranolol hydrochloride	1,3-Dimethyluric acid
8	Guanosine 3',5'-cyclic monophosphate sodium salt (cyclic GMP)	1	G1	1007	[Hydroxy(tosyloxy)iodo]benzene	1-Aminocyclopropanecarboxylic acid
9	N-Succinyl-Ala-Ala-Pro-Phe p-nitroanilide	1	H1	1050	1-(4-Chlorobenzyl)-5-methoxy-2-methylindole-3-acetic acid	1-Methylhistamine dihydrochloride
10	Coenzyme A sodium salt hydrate	1	A2	1009	1,3-Dimethyluric acid	1-Phenyl-1-cyclopropanecarboxylic acid
11	DL-Pantothenic acid hemicalcium salt Pantothenate (vitamin B5)	1	B2	1010	1-Aminocyclopropanecarboxylic acid	1-Phenylcyclopentanecarboxylic acid 2-(2-diethylaminoethoxy)ethyl ester citrate salt
12	b-Nicotinamide adenine dinucleotide phosphate reduced tetrasodium salt (NADP+,NADPH)	1	C2	1011	1-Methylhistamine dihydrochloride	2,3,5-Triiodobenzoic acid
13	b-Nicotinamide adenine dinucleotide (NAD+, NADH)	1	D2	1012	1-Phenyl-1-cyclopropanecarboxylic acid	2-Amino-4-methylphenol
14	Thiamine hydrochloride (Vitamin B1)	1	E2	1013	1-Phenylcyclopentanecarboxylic acid 2-(2-diethylaminoethoxy)ethyl ester citrate salt	2-Amino-5-methylthiazole
15	Thiamine pyrophosphate (TPP)	1	F2	1014	2,3,5-Triiodobenzoic acid	2-Aminofluorene
16	Tetrahydrofolate (Folic acid, vitamin M) Pteroyl-L-glutamic acid	1	G2	1015	2-Amino-4-methylphenol	2-Aminopyridine
17	Cyanocobalamin (Vitamin B12)	1	H2	1008	2-Amino-5-methylthiazole	2-Chlorodimedone
19	Pyridoxine (Vitamin B6)	1	B3	1018	2-Aminofluorene	2'-Deoxyadenosine 3'-monophosphate
20	6,7-Dimethyl-5,6,7,8-tetrahydropterine hydrochloride	1	C3	1048	2-Aminopyridine	2'-Deoxyadenosine 5'-monophosphate
21	Flavin adenine dinucleotide disodium salt dihydrate flavin adenine dinucleotide (FAD, FADH2)	1	D3	1020	2-Chlorodimedone	2-Ethylimidazole
23	Nicotinic acid (Niacin,vitamin B)	1	F3	1022	2'-Deoxyadenosine 3'-monophosphate	2-Methylimidazole
24	D-(+)-Neopterin	1	G3	1051	2'-Deoxyadenosine 5'-monophosphate	3,3',5-Triiodothyroacetic acid
25	6-Hydroxydopamine hydrochloride (6-hydroxyDOPA)	1	H3	1016	2-Ethylimidazole	3,4-Dimethylaniline
26	Sodium glycocholate hydrate (Bile salt)	1	A4	1001	2-Methylimidazole	3,4-Dimethylphenol
27	L-Glutathione reduced	1	B4	1002	3,3',5-Triiodothyroacetic acid	3-Hydroxy-4-methoxyphenethylamine hydrochloride, 4-Methoxytyramine hydrochloride
28	(±)-6-Methyl-5,6,7,8-tetrahydropterine dihydrochloride; (Tetrahydropteridine)	1	C4	1003	3,4-Dimethylaniline	3-Isobutyl-1-methylxanthine
29	Biotin (vitamin H)	1	D4	1004	3,4-Dimethylphenol	3-Methoxytyramine hydrochloride
30	Methoxatin (PQQ)	1	E4	1005	3-Hydroxy-4-methoxyphenethylamine hydrochloride, 4-Methoxytyramine hydrochloride	4-Deoxypyridoxine hydrochloride
31	4-Phenylazobenzyloxycarbonyl-Pro-Leu-Gly-Pro-D-Arg	1	F4	1053	3-Isobutyl-1-methylxanthine	4-Guanidinobutyric acid
32	N-tert-Butyl-a-phenylnitrone	1	G4	1050	3-Methoxytyramine hydrochloride	4-Hydroxytamoxifen
33	PABA, vitamin H1	1	H4	1008	4-Deoxypyridoxine hydrochloride	4-Methylpyrazole hydrochloride
34	L-Carnitine inner salt (vitamin Bt)	1	A5	1009	4-Guanidinobutyric acid	5,5-Diphenylhydantoin
36	S-(5'-Adenosyl)-L-methionine chloride	1	C5	1011	4-Hydroxy-3-methoxyphenylglycol sulfate potassium salt	5a-Androstane-3,17-dione
38	L-Aspartic acid (Asp)	1	D5	1012	4-Hydroxytamoxifen	5-Hydroxyindole-3-acetic acid
39	L-Alanine (Ala)	1	E5	1013	4-Methylpyrazole hydrochloride	Acebutolol hydrochloride
40	L-Leucine (Leu)	1	F5	1006	5,5-Diphenylhydantoin	Acetazolamide
41	L-Isoleucine (Ile)	1	G5	1007	5a-Androstane-3,17-dione	Acetylsalicylic acid (aspirin)
42	L-Tyrosine	1	H5	1016	5-Hydroxyindole-3-acetic acid	Acetylthiocholine chloride
43	L-Tryptophan (Trp)	1	A6	1017	6,7-Dimethyl-5,6,7,8-tetrahydropterine hydrochloride	Acycloguanosine
44	L-Serine (Ser)	1	B6	1018	Acebutolol hydrochloride	Adenosine 3',5'-cyclic monophosphate (cyclic AMP)
45	L-Threonine	1	C6	1019	Acetazolamide	AEBSF
46	L-Glutamic acid (Glu), (S)-(+)-Glutamic acid	1	D6	1020	Acetylsalicylic acid (aspirin)	Aflatoxin B1
47	L-Glutamine	1	E6	1021	Acetylthiocholine chloride	Alaproclate hydrochloride
48	L-Methionine (Met)	1	F6	1022	Acycloguanosine	Amantadine hydrochloride
49	L-Proline (Pro)	1	G6	1023	Adenosine 3',5'-cyclic monophosphate (cyclic AMP)	Ampicillin
50	trans-4-Hydroxy-L-proline	1	H6	1024	AEBSF	Androstenedione
51	L-Arginine (Arg)	1	A7	1009	Aflatoxin B1	Antipyrine
52	L-Lysine (Lys), (S)-(+)-lysine	1	B7	1010	Alaproclate hydrochloride	AY 9944
53	L-Histidine (His)	1	c7	1011	Amantadine hydrochloride	benzamide
54	O-Phospho-L-serine	1	D7	1012	Ampicillin	Benzamidine hydrochloride hydrate
55	O-Phospho-L-tyrosine	1	E7	1013	Androstenedione	Benzoic acid
56	PMSF	1	F7	1014	Antipyrine	Benzylamine
57	AEBSF	1	G7	1015	AY 9944	Bestatin hydrochloride
58	4-Chloromercuribenzoic acid	1	H7	1016	benzamide	Bisbenzimide H 33258
59	Phosphoramidon disodium salt	1	A8	1052	Benzoic acid	Bithionol
60	Pepstatin A	1	B8	1018	Benzylamine	b-Nicotinamide adenine dinucleotide (NAD+, NADH)
61	Daphnetin	1	C8	1019	Bisbenzimide H 33258	Brefeldin A
62	suramin sodium salt	1	D8	1020	Bithionol	Bromocresol Green
63	2,3-diphosphoglycerate pentasodium salt	1	E8	1069	Brefeldin A	Bromophenol Blue
64	Penicillin G potassium salt (benzyl penicillin)	1	F8	1022	Bromocresol Green	Camptothecin
65	Rifampicin	1	G8	1023	Bromophenol Blue	Carbamazepine
66	Puromycin dihydrochloride hydrate	1	H8	1024	Camptothecin	Chelerythrine chloride
67	cycloheximide	1	A9	1001	Carbamazepine	Chelidamic acid
68	cytochalasin B	1	B9	1002	Chelerythrine chloride	Chloramphenicol
69	colchicine	1	C9	1003	Chelidamic acid	Chlorpropamide
70	Acetylsalicylic acid (aspirin)	1	D9	1004	Chloramphenicol	Choline phosphate chloride
71	ouabain	1	E9	1005	Chlorpropamide	Chymostatin
72	neostigmine	1	F9	1006	Choline phosphate chloride	Cimetidine
73	N-Phospho-Ile(O-ethyl)-Tyr(O-benzyl)-Gly dipotassium salt	1	G9	1007	Chymostatin	Cinoxacin
74	MDL 28170 Carbobezoxy-valinyl-phenylalaninal	1	H9	1008	Cimetidine	Clenbuterol hydrochloride
75	Cantharidin	1	A10	1009	Cinoxacin	clofibrate
76	Calmidazolium chloride	1	B10	1010	Clenbuterol hydrochloride	Coenzyme A sodium salt hydrate
77	Sodium oxamate	1	C10	1054	clofibrate	colchicine
78	Chymostatin	1	D10		Coenzyme A sodium salt hydrate	compound
79	Ellipticine	1	E10	1049	colchicine	Cordycepin
80	Aminophylline hydrate	1	F10	1014	Cordycepin	Cromolyn sodium salt
81	erythro-9-(2-Hydroxy-3-nonyl)adenine hydrochloride	1	G10	1015	Cromolyn sodium salt	cycloheximide
82	Ceftriaxone sodium salt	1	H10	1016	Curcumin	Cyproheptadine
83	Cinoxacin	1	A11	1001	cycloheximide	Cytidine 5'-monophosphate
84	Doxycycline hyclate Doxycycline hydrochloride hemiethanolate hemihydrate	1	B11	1002	Cyproheptadine	Cytidine 5'-triphosphate disodium salt (CTP)
85	Nalidixic acid sodium salt	1	C11	1003	Cytidine 5'-monophosphate	Dansylcadaverine
86	Cytosineb-D-arabinofuranoside	1	D11	1053	Cytidine 5'-triphosphate disodium salt (CTP)	Dansylglycin
87	Praziquantel	1	E11	1005	Dansylcadaverine	Daphnetin
88	Ethosuximide	1	F11	1006	Dansylglycin	Diclofenac sodium salt
89	L-Methionine sulfoximine	1	G11	1052	Daphnetin	Didecyldimethylammonium bromide
90	8-Methoxypsoralen	1	H11	1008	Diclofenac sodium salt	Digitoxin
91	2-Chlorodimedone	1	A12	1052	Didecyldimethylammonium bromide	Dipyrone
92	DMCM Methyl-6,7-dimethoxy-4-ethyl-b-carboline-3-carboxylate	1	B12	1018	Digitoxin	DL-Pantothenic acid hemicalcium salt Pantothenate (vitamin B5)
93	Acetamide	1	C12	1011	Dipyrone	DL-Thiorphan
95	4-Aminobenzamidine dihydrochloride	1	E12	1021	DL-Pantothenic acid hemicalcium salt Pantothenate (vitamin B5)	Dobutamine hydrochloride
96	Chelerythrine chloride	1	F12	1051	DL-Thiorphan	erythro-9-(2-Hydroxy-3-nonyl)adenine hydrochloride
97	Benzamidine hydrochloride hydrate	1	G12	1007	Dobutamine hydrochloride	Ethosuximide
98	Diethylenetriaminepentaacetic acid	1	H12	1024	erythro-9-(2-Hydroxy-3-nonyl)adenine hydrochloride	Ethyl 3-pyridineacetate
99	Boc-L-phenylalaninol	2	A1	1049	Ethosuximide	Flutamide
100	4-Methylpyrazole hydrochloride	2	B1	1026	Ethyl 3-pyridineacetate	Furosemide
101	benzamide	2	C1	1070	Flutamide	Griseofulvin
102	Bepridil hydrochloride	2	D1	1020	Furosemide	Guanosine 3',5'-cyclic monophosphate sodium salt (cyclic GMP)
103	Indomethacin morpholinylamide	2	E1	1021	Griseofulvin	Guanosine 5'-triphosphate sodium salt hydrate (GTP)
104	Bay 11-708	2	F1	1022	Guanosine 3',5'-cyclic monophosphate sodium salt (cyclic GMP)	Homovanillic acid
105	Mitoxantrone dihydrochloride	2	G1	1050	Guanosine 5'-triphosphate sodium salt hydrate (GTP)	Hydrochlorothiazide
106	Ribavirin	2	H1	1024	Homovanillic acid	Ibuprofen
107	Allopurinol	2	A2	1017	Hydrochlorothiazide	Idazoxan hydrochloride
108	Lonidamine	2	B2	1026	Ibuprofen	Indomethacin
109	Lorglumide sodium salt	2	C2	1019	Idazoxan hydrochloride	Inosine 5'-monophosphate disodium salt
110	Atropine sulfate salt hydrate	2	D2	1028	Indomethacin	Kynurenic acid
111	N-Phenylanthranilic acid	2	E2	1021	Inosine 5'-monophosphate disodium salt	L-Alanine (Ala)
112	(±)-Thalidomide	2	F2	1030	Kynurenic acid	L-Aspartic acid (Asp)
113	(-)-Perillic acid	2	G2	1070	L-Alanine (Ala)	L-Carnitine inner salt (vitamin Bt)
114	3-Amino-2,3-dihydrobenzoic acid hydrochloride	2	H2	1032	L-Arginine (Arg)	L-Glutathione reduced
115	L-2-Aminoadipic acid	2	A3	1017	L-Carnitine inner salt (vitamin Bt)	L-Leucine (Leu)
116	Chelidamic acid	2	B3	1026	L-Glutathione reduced	L-Leucine-2-naphthylamide
117	Idazoxan hydrochloride	2	C3	1019	Lidocaine hydrochloride	Lorglumide sodium salt
118	3'-Azido-3'-deoxythymidine	2	D3	1028	L-Leucine (Leu)	L-Pipecolic acid
119	Acycloguanosine	2	E3	1106	L-Leucine-2-naphthylamide	L-Tyrosine (Tyr)
120	7,7-Dimethyl-(5Z,8Z)-eicosadienoic acid	2	F3	1030	L-Lysine (Lys), (S)-(+)-lysine	Lumicolchicine
121	Ranolazine dihydrochloride	2	G3	1023	Lonidamine	Mecamylamine hydrochloride
122	[Hydroxy(tosyloxy)iodo]benzene	2	H3	1052	Lorglumide sodium salt	Methiothepin mesylate salt
123	1-(4-Chlorobenzyl)-5-methoxy-2-methylindole-3-acetic acid	2	A4	1047	L-Pipecolic acid	Methotrexate, (+)-Amethopterin
124	a-Methyl-DL-tyrosine	2	B4	1026	L-Tyrosine (Tyr)	Mycophenolic acid
125	6-Methoxy-1,2,3,4-tetrahydro-9H-pyrido[3,4-b]indole	2	C4	1027	Lumicolchicine	N-Acetylneuraminic acid
126	Captopril	2	D4	1028	Mecamylamine hydrochloride	Nalidixic acid sodium salt
127	(±)-2,3-Dichloro-a-methylbenzylamine hydrochloride	2	F4	1030	Methiothepin mesylate salt	Naproxen
128	3,5-Dinitrocatechol	2	G4	1031	Methotrexate, (+)-Amethopterin	neostigmine
129	DL-Thiorphan	2	H4	1068	Mycophenolic acid	Netropsin dihydrochloride hydrate
130	Aztreonam	2	A5	1051	N-Acetylneuraminic acid	Novobiocin sodium salt
131	Carbetapentane citrate salt	2	B5	1034	Nalidixic acid sodium salt	N-Phenylanthranilic acid
132	Naloxone hydrochloride dihydrate	2	C5	1027	Naproxen	N-Phospho-Ile(O-ethyl)-Tyr(O-benzyl)-Gly dipotassium salt
133	(-)-Tetramisole hydrochloride	2	D5	1028	neostigmine	N-Succinyl-Ala-Ala-Pro-Phe p-nitroanilide
134	Chlorzoxazone	2	E5	1029	Netropsin dihydrochloride hydrate	N-tert-Butyl-a-phenylnitrone
135	Genistein	2	F5	1030	Nitrendipine	O-Phospho-L-tyrosine
136	Methyl-6-O-(N-heptylcarbamoyl)-a-D-glucopyranoside	2	G5		Novobiocin sodium salt	Orange II sodium salt
137	Methiothepin mesylate salt	2	H5	1032	N-Phenylanthranilic acid	ouabain
138	Sepiapterin	2	A6	1049	N-Phospho-Ile(O-ethyl)-Tyr(O-benzyl)-Gly dipotassium salt	Oxacillin
139	N-Acetyl-L-tryptophan 3,5-bis(trifluoromethyl)benzyl ester	2	B6	1034	N-Succinyl-Ala-Ala-Pro-Phe p-nitroanilide	p-Aminohippuric acid
140	Picotamide	2	C6	1027	N-tert-Butyl-a-phenylnitrone	Pentoxifylline
141	Thioctic acid Lipoic acid	2	D6	1050	N-Vanillylnonanamide	Pepstatin A
142	5-Bromo-4-chloro-3-indolylb-D-galactopyranoside	2	E6	1048	O-Phospho-L-tyrosine	Phenol Red
143	a-Lactose monohydrate	2	F6	1038	Orange II sodium salt	phenylbutazone
144	D(+)-Maltose monohydrate	2	G6	1031	ouabain	Phenylpyruvic acid
145	sucrose	2	H6	1032	Oxacillin	Picrotin
146	L-Leucine-2-naphthylamide	2	A7	1053	p-Aminohippuric acid	Podophyllotoxin
147	Glucose	2	B7	1034	Penicillin G potassium salt (benzyl penicillin)	Praziquantel
148	4-Deoxypyridoxine hydrochloride	2	C7	1048	Pentoxifylline	Prednisolone
149	D-Sorbitol	2	D7	1036	Pepstatin A	Procaine hydrochloride
150	Melatonin	2	E7	1029	Phenol Red	Psoralen
151	Serotonin hydrochloride	2	F7	1038	phenylbutazone	Puromycin dihydrochloride hydrate
152	(1R,2S)-(-)-Ephedrine	2	G7	1031	Phenylpyruvic acid	Quercetin dihydrate
153	(±)-Norepinephrine (+)-bitartrate salt Noradrenaline	2	H7	1032	Picrotin	rac-Glycerol 3-phosphate disodium salt
154	(±)-Epinephrine	2	A8	1025	Podophyllotoxin	Ranitidine hydrochloride
155	Dopamine hydrochloride	2	B8	1034	Praziquantel	Ranolazine dihydrochloride
156	g-Aminobutyric acid (GABA)	2	C8	1027	Prednisolone	Resveratrol
157	Hydrocortisone (cortisol)	2	D8	1036	Procaine hydrochloride	Ribavirin
158	caffeine	2	E8	1037	Psoralen	Roscovitine
159	progesterone	2	F8	1038	Quercetin dihydrate	Silibinin
160	estradiol	2	G8	1023	rac-Glycerol 3-phosphate disodium salt	Sodium cacodylate trihydrate
161	vasopressin	2	H8	1040	Ranitidine hydrochloride	Sodium creatine phosphate dibasic tetrahydrate
162	acetylcholine	2	A9	1033	Resveratrol	Sodium DL-lactate
163	g-Butyrolactone	2	B9		Ribavirin	Sodium glycocholate hydrate (Bile salt)
164	Chlorpropamide	2	C9	1035	Roscovitine	Sodium salicylate
165	3-Aminopropionitrile fumarate salt	2	D9	1036	Silibinin	succinate
166	Phenylbutazone	2	E9	1037	Sodium cacodylate trihydrate	Sulfaphenazole
167	Betaine hydrochloride Trimethyl glycine	2	F9	1038	Sodium citrate tribasic dihydrate	suramin
168	Choline bromide	2	G9	1031	Sodium creatine phosphate dibasic tetrahydrate	tert-Butyl carbazate
169	(-)-Cotinine	2	H9	1040	Sodium DL-lactate	Theophylline
170	1-Methylhistamine dihydrochloride	2	A10	1068	Sodium glycocholate hydrate (Bile salt)	Thiamine hydrochloride (Vitamin B1)
171	Sodium creatine phosphate dibasic tetrahydrate	2	B10	1047	Sodium salicylate	Thiamine pyrophosphate (TPP)
172	Lipoic acid	2	C10	1035	succinate	Thymidine 5'-monphosphate
173	(±)-a-Lipoamide	2	D10	1036	Sulfaphenazole	Timolol maleate salt
175	N-Acetyl-L-tryptophanamide	2	F10	1044	tert-Butyl carbazate	Tolbutamide
176	Sodium citrate tribasic dihydrate	2	G10	1039	Theophylline	Tubercidin
177	succinate	2	H10	1040	Thiamine hydrochloride (Vitamin B1)	Tyrphostin 1
178	rac-Glycerol 3-phosphate disodium salt	2	A11	1025	Thiamine pyrophosphate (TPP)	Warfarin
179	Phospho(enol)pyruvate potassium salt	2	B11	1041	Thymidine 5'-monphosphate	

molID	Aor	Bam	Bli	Iba	Hvu	Nuc	DnaG	PrgI	Sav1430	YtfP	PA1324	YkvR	YkfF	STM1790	DGKA	YjbR	UFC1
2	L-Leucineb-naphthylamide	L-Leucineb-naphthylamide	5-Bromo-4-chloro-3-indolylb-D-galactopyranoside	Sodium glycocholate hydrate (Bile salt)	(±)-6-Methyl-5,6,7,8-tetrahydropterine dihydrochlorid; (Tetrahydropteridine)	4-Hydroxy-3-methoxyphenylglycol sulfate potassium salt	1-Methylimidazole	1-Methylimidazole	5-Fluoro-5'-deoxyuridine	Adenine (Vitamine B4)	[Hydroxy(tosyloxy)iodo]benzene	3,5-Dinitrocatechol	Amoxicillin	Pepstatin A	Guanosine 3',5'-cyclic monophosphate sodium salt (cyclic GMP)	Cytidine 5'-triphosphate disodium salt (CTP)	L-Histidine (His)
3	N-Acetylprocainamide hydrochloride	5-Bromo-4-chloro-3-indolylb-D-galactopyranoside	(-)-Perillic acid	Doxycycline hyclate	L-Leucineb-naphthylamide	Acebutolol hydrochloride	(±)-a-Lipoamide	Didecyldimethylammonium bromide	(±)-Thalidomide	Amoxicillin	2',3'-Dideoxyadenosine	Adenine (Vitamine B4)	Cyclo(His-Pro)	Methyl 6,7-dimethoxy-4-ethyl-b-carboline-3-carboxylate	Adenine (Vitamine B4, 6-Aminopurine)	Doxycycline hyclate	suramin
4	1,3-Dimethyluric acid	2-Amino-4-methylphenol	1-Phenyl-1-cyclopropanecarboxylic acid	β-Nicotinamide adenine dinucleotide (NAD+, NADH)	O-Phospho-L-tyrosine	Adenosine 3',5'-cyclic monophosphate (cyclic AMP)	(±)-Propranolol hydrochloride	L-Carnitine inner salt (vitamin Bt)	2',3'-Dideoxyadenosine	Diminazene aceturate (Berenil)	2-Aminothiazol	Amoxicillin	Ethacridine	Flavin adenine dinucleotide disodium salt dihydrate flavin adenine dinucleotide (FAD, FADH2)	Adenosine 5'-monophosphoric acid monohydrate	Nalidixic acid sodium salt	Aminophylline hydrate
5	5-Bromo-4-chloro-3-indolylb-D-galactopyranoside	2'-Deoxyadenosine 5'-monophosphate	2'-Deoxyadenosine 5'-monophosphate	Flavin adenine dinucleotide disodium salt dihydrate flavin adenine dinucleotide (FAD, FADH2)	2-Amino-4-methylphenol	Adenosine 5'-triphosphate disodium salt (ATP)	3-Aminopropionitrile fumarate salt	Methiothepin mesylate salt	2'-Deoxyadenosine 5'-monophosphate	Eserine (Physostigmine)	2-Ethylimidazole	histamine	L-2-Aminoadipic acid	L-Glutamic acid (Glu)	Novobiocin sodium salt *IN FREEZER	suramin	Rifampicin
6	(-)-Perillic acid	3-Methyladenine	4-Deoxypyridoxine hydrochloride	O-Phospho-L-tyrosine	2'-Deoxyadenosine 5'-monophosphate	Aquocobalamin	5,5-Diphenylhydantoin	sucrose	3-(1-Naphthyl)-D-alanine	Kaempferol	4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride	L-2-Aminoadipic acid	Lidocaine hydrochloride	1-Methylhistamine dihydrochloride	Acridine Orange base	Adenosine 3',5'-cyclic monophosphate (cyclic AMP)	4-Methylpyrazole hydrochloride
7	1-Phenyl-1-cyclopropanecarboxylic acid	4-Deoxypyridoxine hydrochloride	5-Phenylvaleric acid	Penicillin G potassium salt (benzyl penicillin)	3,5-Dinitrocatechol	Bepridil hydrochloride	Acycloguanosine		4-Chloromercuribenzoic acid	L-2-Aminoadipic acid	4-Hydroxytamoxifen	Sodium citrate tribasic dihydrate	Lorglumide sodium salt	Acycloguanosine	6,9-Diamino-2-ethoxyacridine-DL-lactate monohydrate	Thiamine pyrophosphate (TPP)	Bay 11-7085
8	2'-Deoxyadenosine 5'-monophosphate	4-Hydroxy-3-methoxyphenylglycol sulfate potassium salt	6,9-Diamino-2-ethoxyacridine-DL-lactate monohydrate	Rifampicin	4-Methylpyrazole hydrochloride	Ciprofloxacin	Chelerythrine chloride		6,7-Dimethyl-5,6,7,8-tetrahydropterine hydrochloride	Lidocaine hydrochloride	5a-Androstane-3,17-dione			Bay 11-7085	Orange II sodium salt	1-Methylhistamine dihydrochloride	3,5-Dinitrocatechol
9	4-Hydroxy-3-methoxyphenylglycol sulfate potassium salt	6,9-Diamino-2-ethoxyacridine-DL-lactate monohydrate	6-Methoxy-1,2,3,4-tetrahydro-9H-pyrido[3,4-b]indole	N-Succinyl-Ala-Ala-Pro-Phe p-nitroanilide	Acridine Orange base	Cytidine 5'-triphosphate disodium salt (CTP)	Didecyldimethylammonium bromide		6-Amino-3-methylpurine	N-Acetyl-L-tryptophan_3-5-trifluoromethyl_benzyl_ester	a-Methyl-DL-tyrosine			Bromophenol Blue	3-Chlorophenol	Bepridil hydrochloride	Acridine Orange base
10	4-Methylpyrazole hydrochloride	Acridine Orange base	Acridine Orange base	L-Arginine (Arg)	Adenosine 5'-monophosphoric acid monohydrate	Diminazene aceturate	L-Histidine (His)		6-Hydroxydopamine hydrochloride (6-hydroxyDOPA)	Novobiocin sodium salt	Bay 11-7085			Uridine 5'-monophosphate	Prednisolone	5-Bromo-4-chloro-3-indolylb-D-galactopyranoside	Bromophenol Blue
11	5-Phenylvaleric acid	Adenosine 5'-monophosphoric acid monohydrate	Bisbenzimide H 33258	Phospho(enol)pyruvate potassium salt	α-Methyl-DL-tyrosine	Guanosine 3',5'-cyclic monophosphate sodium salt (cyclic GMP)	mitoxantrone dihydrochloride		Adenosine 3',5'-cyclic monophosphate (cyclic AMP)	Orange II sodium salt	Bromophenol Blue			Bisbenzimide H 33258	1,3-Dimethyluric acid	histamine	Adenine (Vitamine B4, 6-Aminopurine)
12	6,9-Diamino-2-ethoxyacridine-DL-lactate monohydrate	Adenosine 5'-triphosphate disodium salt (ATP)	Bromophenol Blue	Carbetapentane citrate salt	Bisbenzimide H 33258	Guanosine 5'-triphosphate sodium salt hydrate (GTP)	Sodium creatine phosphate dibasic tetrahydrate		a-Methyl-DL-tyrosine	Phenol Red	Carbobezoxy-valinyl-phenylalaninal			Didecyldimethylammonium bromide	Chlortetracycline hydrochloride	Bromophenol Blue	Hydralazine hydrochloride
13	Acebutolol hydrochloride	Bisbenzimide H 33258	Carbetapentane citrate salt	5-Bromo-4-chloro-3-indolylb-D-galactopyranoside	Bromophenol Blue	L-Leucine (Leu)	Sodium DL-lactate		Biotin (vitamin H)	progesterone	Chlorpropamide			Resveratrol	6-Phosphogluconic acid trisodium salt	N-Propyl-1,3-propanediamine	2',3'-Dideoxyadenosine
14	Acridine Orange base	β-Nicotinamide adenine dinucleotide (NAD+, NADH)	Chlortetracycline hydrochloride	N-Acetyl-L-tryptophanamide	Ciprofloxacin	Lumicolchicine			b-Nicotinamide adenine dinucleotide (NAD+, NADH)	Tetracycline	Diminazene aceturate			Tyrphostin 1	3-Hydroxy-DL-kynurenine	Adenine (Vitamine B4, 6-Aminopurine)	2-Methylimidazole
15	Adenosine 5'-monophosphoric acid monohydrate	Bromophenol Blue	Ciprofloxacin	Methiothepin mesylate salt	Cyclo(His-Pro)	mitoxantrone dihydrochloride			b-Nicotinamide adenine dinucleotide phosphate reduced tetrasodium salt		Ellipticine			1-Octanol	7-Deazaguanine	Clofibrate	Menadione (2-Methyl-1,4-naphthoquinone, vitamin k3)
16	Adenosine 5'-triphosphate disodium salt (ATP)	Carbetapentane citrate salt	Dansylcadaverine	2,3,5,6, Tetramethyl-P-Benzoquinone (Duroquinone)	Dansylcadaverine	Phosphocholine chloride calcium salt tetrahydrate			Bromophenol Blue		Menadione (vitamin k3)			Aquocobalamin	trans-4-(Aminomethyl)cyclohexanecarboxylic acid	Menadione (2-Methyl-1,4-naphthoquinone, vitamin k3)	Didecyldimethylammonium bromide
17	α-Methyl-DL-tyrosine	Ciprofloxacin	Dansylglycin	Bromophenol Blue	Didecyldimethylammonium bromide	suramin			Carbetapentane citrate salt		Nicotinic acid (Niacin,vitamin B)			Ebselen	Dipyrone	3-(1-Naphthyl)-D-alanine	1-Methylimidazole
19	Bisbenzimide H 33258	Cytidine 5'-triphosphate disodium salt (CTP)	Deoxyadenosyl cobalamin (Vitamin B12)	Dansylcadaverine	DL-Pantothenic acid hemicalcium salt Pantothenate (vitamin B5)	Thymidine 5'-triphosphate sodium salt (TTP)			Cytidine 5'-monophosphate		Pepstatin A			2,6-Diisopropylphenol - LIQUID		Flutamide	2-Aminopyridine
20	β-Nicotinamide adenine dinucleotide (NAD+, NADH)	Dansylcadaverine	Didecyldimethylammonium bromide	Thymidine 5'-monphosphate	L-Lysine (Lys)	Uridine 5'-triphosphate trisodium salt dihydrate (UTP)			Dobutamine hydrochloride		PMSF, Phenylmethanesulfonyl fluoride			Nadolol		Diminazene aceturate	Resveratrol
21	Bromophenol Blue	Deoxyadenosyl cobalamin (Vitamin B12)	Flavin adenine dinucleotide disodium salt dihydrate flavin adenine dinucleotide (FAD, FADH2)	Pyridoxal 5'-phosphate	Lorglumide sodium salt				Ethacridine		suramin					PTH-tryptophan	Aquocobalamin
23	Carbetapentane citrate salt	Didecyldimethylammonium bromide	Guanosine 5'-triphosphate sodium salt hydrate (GTP)	2-Amino-4-methylphenol	Lumicolchicine				Genistein		Tyrphostin 25					Lumicolchicine	
24	Ciprofloxacin	Diminazene aceturate	Hydrocortisone (cortisol)	Bisbenzimide H 33258	Methotrexate hydrate				L-Lysine (Lys), (S)-(+)-lysine							(±)-Camphor	
25	Clofibrate	Flavin adenine dinucleotide disodium salt dihydrate flavin adenine dinucleotide (FAD, FADH2)	L-Glutamine (Gln)	2'-Deoxyadenosine 5'-monophosphate	N-Acetyl-L-tryptophanamide				Lonidamine							Homovanillic acid	
26	Cytidine 5'-triphosphate disodium salt (CTP)	Guanosine 3',5'-cyclic monophosphate sodium salt (cyclic GMP)	L-Glutathione reduced	N-Acetylneuraminic acid	N-Vanillylnonanamide				Lorglumide sodium salt							Cephalexin hydrate	
27	Dansylcadaverine	Guanosine 5'-triphosphate sodium salt hydrate (GTP)	N-Acetyl-L-tryptophanamide	Phosphocholine chloride calcium salt tetrahydrate	Phosphocholine chloride calcium salt tetrahydrate				Methiothepin mesylate salt							1-Octanol	
28	Dansylglycin	Hydrocortisone (cortisol)	N-tert-Butyl-a-phenylnitrone	Ciprofloxacin	Pyridoxal 5'-phosphate				Novobiocin sodium salt							(±)-Verapamil hydrochloride	
29	Deoxyadenosyl cobalamin (Vitamin B12)	Hydrocortisone (cortisol)	Penicillin G potassium salt (benzyl penicillin)	Diminazene aceturate	S-(-)-Carbidopa				O-Phospho-L-tyrosine							Tetracycline	
30	Didecyldimethylammonium bromide	Methotrexate hydrate	Pentoxifylline	4-Hydroxy-3-methoxyphenylglycol sulfate potassium salt	Thymidine 5'-monphosphate				ouabain							2,6-Diisopropylphenol - LIQUID	
31	D-Panthenol (R)-(+)-2,4-Dihydroxy-N-(3-hydroxypropyl)-3,3-dimethylbutyramide) *IN	mitoxantrone dihydrochloride	Phosphocholine chloride calcium salt tetrahydrate	Lumicolchicine	Uridine 5'-monophosphate				Praziquantel							Bithionol	
32	Eserine (Physostigmine)	N-tert-Butyldimethylsilyl-N-methyltrifluoroacetamide		Cimetidine	Uridine 5'-triphosphate trisodium salt dihydrate (UTP)				Resorufin								
33	Flavin adenine dinucleotide disodium salt dihydrate flavin adenine dinucleotide (FAD, FADH2)	Phosphocholine chloride calcium salt tetrahydrate							Sodium palmitate								
34	Hydrocortisone (cortisol)	Pyridoxal 5'-phosphate															
36	Menadione (2-Methyl-1,4-naphthoquinone, vitamin K3)	S-(-)-Carbidopa *IN SUBZERO															
38	mitoxantrone dihydrochloride	Sodium cacodylate trihydrate															
39	Nalidixic acid sodium salt	Thymidine 5'-monphosphate															
40	Phosphocholine chloride calcium salt tetrahydrate	Thymidine 5'-triphosphate sodium salt (TTP)															
41	Thymidine 5'-monphosphate																

180 Sodium Di-lactate
181 Sodium pyruvate
182 3-Indoleacetic acid
183 Lithium acetoacetate
184 Histamine
185 porphobilinogen
186 Acetylthiocholine chloride
187 N-Vanillylnonanamide
192 (±)-Carnitine chloride
193 Dantrolene
194 Pentyl vinyl carbinol
195 Adenine (Vitamin B4)
196 Nimesulene (vitamin h2)
197 D-Panthenol
198 3,5-Diiodo-L-tyrosine dihydrate
199 N,O-Naphthoyl-D-alanine
200 Theophylline
201 Alaproclate hydrochloride
202 Guanidoh dihydrate
203 Hydralazine hydrochloride
204 Exemie (Physostigmine)
205 Neodecin sodium salt
206 Dibenzyldimethylammonium bromide
207 S-(-)-Carbidopa
208 N-p-Tosyl-L-phenylalanine chloromethyl ketone
209 3-Isobutyl-1-methylxanthine
210 Lithium potassium acetyl phosphate
211 12-Hydroxydodecanoic acid
212 D-(-)-Quinate acid lactone
213 Nalidixic acid
214 3-Iodo-L-tyrosine
215 Cytidine PhD
216 1-Aminocyclopropanecarboxylic acid
217 3-Indoleacetic acid
218 Gemcitabienine
219 2-Fluoroaniline
220 Tuberidin
221 SV-(-)-Galactosamine hydrochloride
222 1-Methylimidazole
224 Phenylacrylonic acid
225 Pentaerythrine glycol
226 N-Propyl-1,3-propanediamine
227 2'-3'-Dideoxyadenosine
228 Pyridoxal 5'-phosphate
229 2-Aminothiazol
230 2-Hydroxyethyl disulfide
231 Androsterone
233 Umbilicaccinic acid
234 Dimethyl 2-oxoglutarate
235 D-Cycline
236 2-Cyclohexen-1-one
237 Pseudothiohydantoin
238 Cyclohexylamine
239 Floridin
240 (-)-Bilobalide Ginkgo biloba leaves
242 Tyramine
243 Triethylenetetramine Trientine
244 2-Aminofluorene
245 Ethidium bromide
290 2-Amino-4-methylthiazole
291 2-Amino-5-methylthiazole
292 2-Amino-4-methylphenol, 2-Amino-p-cresol
293 2-Aminophenol
294 2'-Deoxyadenosine 5'-monophosphate
295 2-Aminopyridine
297 2-Ethylimidazole
298 3,4-Dimethylphenol
299 (6-p-Chlorophenyl)alanine
300 2-Methylimidazole
301 2-Methylglutaric acid
302 3,4-Dimethylaniline
303 Carbutapin
304 3-Aminopyridine
305 3-Chlorophenol
306 Adenosine 5'-monophosphoric acid
307 Cytidine 5'-monophosphate
308 2'-Deoxyguanosine 5'-monophosphate
309 Inosine 5'-monophosphate disodium salt
310 Thymidine 5'-monophosphate
311 Uridine 5'-monophosphate
312 Kaempferol
313 Bisbenzimide H 33258
314 Nemopon dihydrochloride hydrate
315 Choline phosphate chloride
316 Warfarin
317 5,7-Dihemethylchromon
318 2'-Deoxyadenosine 5'-monophosphate
320 Chlorobactocycline hydrochloride
321 Buspirone hydrochloride
322 Clenbuterol hydrochloride
323 Aconine Orange base
324 3-Guanidinecarboxylic acid
325 Ciprofloxacin
326 6,9-Diamino-2-ethoxyacridine-DL-lactate monohydrate (Ethacridine)
327 Abbotin hollow-R
328 Orange II sodium salt
330 Aldosterone
331 Ampicillin
332 Furosemide
333 Ibuprofen e-Methyl-4-(isobutylphenyl)acetic acid 2-(4-ISOBUTYLPHENYL)PROPIONIC ACID
335 Aquacobalamin
337 Digitoxin
338 Oxacillin
339 Phenol Red
340 Bromocresol Green
341 Bromophenol Blue
343 Naproxen
344 Clofibrate
345 Sodium salicylate
346 4-Amino-3-methylpurine
348 (+)-Pseudoephedrine hydrochloride
350 N-Acetylneuraminic acid
351 Methyl 4-hydroxybenzoate
352 3,3',5-Triiodothyronacetic acid
353 Desmosterin
354 3-Phenylvalenic acid
355 Pfm tryptophan
356 4-Phosphogluconic acid trisodium salt (3-Gluconate 6-phosphate
357 N-(2H-Benzotriazol-1-ylphenylmethyl)benzamide
358 7-Deoxyguanine
359 N-Acetyl-D-glucosamine
360 Benzylamine
361 Deoxyglycin
362 Agmatine sulfate salt
363 Sa Androstane-3,17-dione
364 Nicorufin acetate
365 Cephalexin hydrate
366 trans-4-(Aminomethyl)cyclohexanecarboxylic acid Tranexamic acid
367 Androstenedione 4-Androstene-3,17-dione
368 Acetazolamide
369 Benzoin hydrochloride
370 Diminazene aceturate Berenil
371 Sodium cevadylate trihydrate
372 (+)-Camphor
373 Cefadin
374 Chloramphenicol
375 Cinderocin
376 Methotrexate hydrate, (+)-Amethopterin
377 Amoxicillin
378 (-)-Ardigenin
380 Oxalinic acid
381 Daunorubicin
382 5-Azacytidine
383 3-Iodobenzyl-N
384 Doxorubicin hydrochloride, Adriamycin, Hydroxydaunorubicin
386 Podophyllotoxin
387 Lumicolchicine
389 N,N'-Dicyano-2,5-dimethylbenzoquinone-diimine
390 Glycerol
391 Salinbac Saffold
392 Diethylcitrillenol
393 N-7-Ethylcytochrome
394 Roscovitine
395 Tyrphostin 1
396 Tyrphostin 25
397 Curcumin
398 5-Fluoro-5'-deoxyuridine
399 Tuaftalin
401 Resveratrol
402 Gitoxin
403 Cinnolizine
404 Cyclophosphamide monohydrate
405 5-Octanol
406 Tolbutamide
410 tert-Butyl carbazate
411 AY 9944
412 Amantadine hydrochloride
413 p-Aminohippuric acid
416 Psoralen
417 5-Hydroxyindole-3-acetic acid
418 4-Hydroxy-3-methoxyphenylglycol sulfate potassium salt
419 N-tert-Butyldimethylsilyl-N-methyltrifluoroacetamide
420 Adenosylhomo
421 Affetosan B1
1035 Timolol maleate salt
1042 Tolbutamide
1037 Tuberidin
1044 Tryptophan 1
1039 Uridine 5'-monophosphate
1040 Warfarin
1041
1042
1043
1044
1045
1046
1033
1071
1069
1033
1058
1046
1033
1074
1043
1044
1045
1074
1056
1056
1069
1058
1068
1069
1055
1067
1067
1057
1070
1055
1066
1067
1070
1055
1055
1090
1067
1111
1076
1079
1082
1085
1087
1092
1033
1076
1079
1082
1044
1095
1095
1085
1082
1088
1090
1093
1095
1076
1105
1092
1088
1085
1077
1083
1080
1083
1086
1088
1091
1080
1077
1091
1086
1089
1091

422	DL-4-Hydroxy-3-methoxymandelic acid	4 G9	1094
423	Tamoxifen	4 H8	1096
424	Diclofenac sodium salt	4 H10	1078
425	Benoxic acid	4 B10	1080
426	Ethylenediaminetetrasoratic acid, Edathamil	4 C10	1084
427	2-Fluoro-2-deoxy-D-glucose 6-phosphate barium salt	4 D10	1086
428	Acetrizoate	4 E10	1089
429	Bekfeldin A	4 F10	1110
431	1,7-Dimethylxanthine	4 H10	1113
432	Esmlol	4 A11	1078
433	Flutamide	4 B11	1081
434	4-Guanidinobutyric acid	4 C11	1097
435	Haloperidol-monydrate I	4 D11	1087
436	Homovanillic acid	4 E11	1089
437	3-Hydroxy-DL-kynurenine	4 F11	1092
438	3-Hydroxy-4-methoxyphenethylamine hydrochloride, 4-Methoxytyramine hydrochloride	4 G11	1094
439	4-Hydroxycoumarin	4 H11	1096
440	Leflunomide	4 A12	1100
441	3-Methoxytyramine hydrochloride	4 B12	1106
443	Minoxidil	4 E12	1108
444	Motiobarbine	4 F12	1111
445	Myristoyl-from Acyrotu stevita	4 G12	1111
446	Dobutamine hydrochloride	5 B1	1105
449	Pentadiflibe	5 C1	1105
450	Amitriptyline hydrochloride	5 D1	1104
452	Bithionol	5 F1	1107
454	Griseofulvin	5 H1	1107
456	Indomethacin	5 A2	1103
457	Kynurenic acid	5 B2	1081
458	Myristphenolic acid	5 C2	1084
459	Nitrendipine	5 D2	1087
460	(+)-Pulpomidol hydrochloride	5 E2	1089
461	Mifedipine	5 G2	1094
462	N-Acetylprocainamide hydrochloride, Azocainide hydrochloride	5 H2	1096
463	1-Pyrrolic acid	5 A3	1113
464	1,3-Dimethyluric acid	5 B3	1081
465	Camoparidol	5 C3	1084
466	4-Hydroxycoumarin	5 D3	
467	Cyproheptadine	5 E3	1090
468	(-)-Bucuzidine methiodide	5 F3	1092
469	Miracurcumine hydrochloride	5 G3	1094
470	Metergoline	5 H3	1066
471	(±)-Verapamil hydrochloride	5 A4	1078
475	3-Phenyl-1-cyclopropanecarboxylic acid	5 G4	1112
476	2-Pyridinecetic acid hydrochloride	4 B5	1112
477	Ethyl 3-pyridineacetate	4 E5	1113
478	2-Acetamidophenol	4 F6	1110
482	Dimethyl 4-methoxyisophthalate	5 A3	1113
483	5-Aminosalicylic acid, Mesalamine, 5-AS	5 D4	1112
484	2-Methylhexol	5 E4	
485	Myristic acid	4 B12	1100
486	2,3,5-Triiodobenzoic acid	5 F2	1105
487	2,6-Diisopropylphenol	5 G1	1097
488	Halothane	5 F4	
489	Azidothymidine	5 G4	
490	Hyoshiguanosine	5 H4	
491	Cefuroxime sodium salt	5 A5	1101
492	5-Fluorocytosine	5 B5	1107
493	Cromolyn sodium salt	5 C5	1104
494	Etoxtein	5 D5	1110
495	Methotrexate, (+)-Amethopterin	5 E5	1098
496	Antipyrine	5 F5	1101
497	Atenolol	5 G5	1098
498	Chloramphenicol	1 c7	
499	Cinnidrine	5 A6	1101
500	Chlorpropamide	5 B6	1107
501	Hydrochlorothiazide	5 C6	1109
502	Tolbutamide	5 D6	
503	Nadolol	5 E6	1098
504	Prednisolone 1,4-Pregnadiene-11B,17a,21-triol-3,20-dione 1-Dehydrocortisol 1-Dehydrohydrocortisone	5 F6	1099
505	(-)-Scopolamine hydrobromide trihydrate Hyoscine hydrobromide Scopine tropate	5 G6	1100
506	Timolol maleate salt	5 H6	1099
507	(±)-Metoprolol (+)-tartrate salt	5 A7	1105
508	Lidocaine hydrochloride	5 B7	1108
509	Sulfaphenazole	5 C7	1109
510	Acebutolol hydrochloride	5 D7	1111
511	Procaine hydrochloride	5 E7	1098
512	Clonidine hydrochloride	5 F7	1099
513	Ranitidine hydrochloride	5 G7	1095
514	Camptothecin	5 H7	1100
515	Tetracycline	5 A8	1102
517	5-Hydroxyisophthalic acid	4 G8	1093
518	Carbamazepine	4 E10	1101

*Only the compound shown to bind a protein are listed. The entire list of compounds comprising the functional screening library is listed in the first column

Supplementary Tables for "A Correlation between Protein Function and Ligand Binding Profiles" Matthew D. Shortridge, Michael Bokemper, Jennifer C. Copeland, Jamie Stark and Robert Powers*

Supplementary Table 2S: Functional similarity scores measured by FunSimMat between all pairs of proteins.[a]

UniProt ID		HSA	BSA	Primase_SA	PrgI	SAV1430	YtfP	PA1324	YkvR	YkfF	Nuclease	Aor	Hvu	Bam	Bli	Iba	STM1790	DGKA	YjbR	UFC1
P02768	HSA	0.98	0.98	0.07	0.28	-	-	-	-	-	-	0.05	0.02	0.04	0.04	0.03	-	0.23	-	0.49
P02769	BSA	0.98	0.98	0.07	0.28	-	-	-	-	-	-	0.05	0.03	0.04	0.04	0.04	-	0.25	-	0.49
Q5HFJ8	Primase_SA	0.07	0.07	0.98		-	-	-	-	-	-	0.2	0.15	0.19	0.19	0.17	-	0.24	-	0.15
P41784	PrgI	0.28	0.28		0.99	-	-	-	-	-	-	0	0	0	0	0	-	0	-	0
Q99U58	SAV1430	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P0AE50	YtfP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Q9I420	PA1324	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
O31683	YkvR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P75677	YkfF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P00644	Nuclease	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P0C1B3	Aor	0.05	0.05	0.2	0	-	-	-	-	-	-	0.74	0.64	0.68	0.68	0.67	-	-	-	-
P16098	Hvu	0.02	0.03	0.15	0	-	-	-	-	-	-	0.64	0.72	0.63	0.63	0.71	-	-	-	-
P00692	Bam	0.04	0.04	0.19	0	-	-	-	-	-	-	0.68	0.63	0.68	0.68	0.63	-	-	-	-
P06278	Bli	0.04	0.04	0.19	0	-	-	-	-	-	-	0.68	0.63	0.68	0.68	0.63	-	-	-	-
P10537	Iba	0.03	0.04	0.17	0	-	-	-	-	-	-	0.67	0.71	0.63	0.63	0.71		0.07	-	0.22
Q8ZP25	STM1790	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P23743	DGKA	0.23	0.25	0.24	0	-	-	-	-	-	-	0.14	0.07	0.11	0.11	0.07	-	0.96	-	0.03
P0AF51	YjbR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Q9Y3C8	UFC1	0.49	0.49	0.15	0	-	-	-	-	-	-	0.2	0.22	0.07	0.07	0.22	-	0.03	-	0.98

[a]The absence of a functional similarity score indicates that at least one of the two proteins lacked GO annotation terms.

Supplementary Tables for "A Correlation between Protein Function and Ligand Binding Profiles" Matthew D. Shortridge, Michael Bokemper, Jennifer C. Copeland, Jamie Stark and Robert Powers*

Supplementary Table 3S: Pairwise comparison of all ligand binding profiles used in this study (E-values).[a]

	HSA	BSA	Primase_SA	PrgI	SAV1430	YtfP	PA1324	YkvR	YkfF	Nuclease	Aor	Hvu	Bam	Bli	Iba	STM1790	DGKA	YjbR	UFC1
HSA	2.32E-68	2.16E-58	5.16E-02	6.48E-02	2.38E-02	1.29E+00	1.19E+00	-	-	2.18E-02	6.02E-02	2.43E-03	4.03E-01	2.73E-02	1.88E-02	4.89E-02	4.23E+01	2.22E-02	1.58E-01
BSA	2.16E-58	1.32E-68	1.35E-02	5.98E-02	3.25E-02	7.08E+00	3.17E-03	-	-	2.82E-04	5.89E-04	6.43E-01	7.33E-02	3.48E-02	3.59E-02	2.82E-04	3.22E+01	3.05E-02	2.22E-02
Primase_SA	5.16E-02	1.35E-02	2.09E-18	8.18E-03	-	-	-	-	-	1.26E-01	1.25E-01	3.30E-02	1.25E-01	4.11E-02	-	6.92E-02	-	-	1.08E-02
PrgI	6.48E-02	5.98E-02	8.18E-03	9.58E-10	1.48E-01	-	-	-	-	-	1.42E-01	1.48E-01	1.44E-01	1.48E-01	1.48E-01	1.30E-01	-	-	1.88E-02
SAV1430	2.38E-02	3.25E-02	-	1.48E-01	3.27E-34	1.03E-02	8.46E-02	-	-	6.86E-02	3.77E-02	6.19E-02	8.30E-02	1.02E-01	1.81E-02	6.86E-02	-	2.80E-03	1.03E-01
YtfP	1.29E+00	7.08E+00	-	-	1.03E-02	1.75E-19	-	1.32E-02	-	-	5.39E-03	-	-	-	-	-	1.31E-01	2.32E-02	-
PA1324	5.43E-01	3.17E-03	-	-	8.46E-02	-	2.96E-26	-	-	1.17E-01	4.21E-02	9.97E-02	5.08E-02	1.02E-01	9.97E-02	3.27E-02	-	7.67E-02	5.47E-03
YkvR	5.10E+00	-	-	-	-	5.59E-04	-	3.91E-11	-	-	-	-	-	-	-	-	-	1.46E-01	1.43E-01
YkfF	1.47E-01	4.27E+00	-	-	5.35E-02	5.59E-04	-	3.12E-03	3.91E-11	-	-	5.08E-02	-	-	-	-	-	-	-
Nuclease	2.18E-02	2.82E-04	1.26E-01	-	6.86E-02	-	1.17E-01	-	-	1.93E-24	8.32E-05	1.56E-02	2.84E-08	6.46E-02	2.25E-03	4.96E-02	7.66E-02	1.97E-04	1.13E-01
Aor	6.02E-02	5.89E-04	1.25E-01	1.42E-01	3.77E-02	5.39E-03	4.21E-02	-	-	8.32E-05	2.83E-38	2.98E-08	6.38E-19	1.19E-15	2.98E-08	2.99E-02	2.12E-02	7.34E-03	7.09E-03
Hvu	2.43E-03	6.43E-01	3.30E-02	1.48E-01	6.19E-02	-	9.97E-02	-	-	1.56E-02	2.36E-09	6.69E-32	1.17E-10	3.86E-06	2.45E-09	1.56E-02	1.24E-01	8.69E-03	2.85E-03
Bam	4.03E-01	7.33E-02	1.25E-01	1.44E-01	8.30E-02	-	5.08E-02	-	-	2.84E-08	3.18E-20	1.66E-09	1.25E-37	1.42E-14	7.56E-12	2.76E-02	1.95E-02	7.63E-02	9.51E-02
Bli	2.73E-02	3.48E-02	4.11E-02	1.48E-01	1.02E-01	-	1.02E-01	-	-	6.46E-02	1.19E-15	3.86E-06	2.62E-13	9.48E-33	2.43E-08	1.40E-02	5.19E-02	1.13E+00	7.08E-02
Iba	1.88E-02	3.59E-02	-	1.48E-01	1.81E-02	-	9.97E-02	-	-	2.25E-03	2.98E-08	2.45E-09	7.56E-12	2.43E-08	1.73E-33	6.86E-02	-	1.40E-02	1.09E-01
STM1790	4.89E-02	2.82E-04	6.92E-02	1.30E-01	6.86E-02	-	3.27E-02	-	-	4.96E-02	2.99E-02	1.56E-02	2.76E-02	1.40E-02	6.86E-02	1.93E-24	-	1.40E-02	3.34E-04
DGKA	4.23E+01	3.22E+01	-	-	-	1.31E-01	-	-	-	7.66E-02	2.12E-02	1.24E-01	1.95E-02	5.19E-02	-	-	1.55E-22	6.50E-04	1.03E-01
YjbR	2.22E-02	3.05E-02	-	-	2.80E-03	2.32E-02	7.67E-02	-	-	1.97E-04	7.34E-03	8.69E-03	7.63E-02	1.71E-02	1.40E-02	1.40E-02	6.50E-04	3.79E-31	1.66E-02
UFC1	1.58E-01	2.22E-02	1.08E-02	1.88E-02	1.03E-01	-	5.47E-03	-	-	1.13E-01	7.09E-03	2.85E-03	9.51E-02	7.08E-02	1.09E-01	3.34E-04	1.03E-01	1.66E-02	2.33E-25

[a]The absence of an E-value indicates the two proteins did not have any binding ligands in common.