



Published in final edited form as:

J Chem Inf Model. 2011 February 28; 51(2): 408–419. doi:10.1021/ci100369f.

A machine learning based method to improve docking scoring functions and its application to drug repurposing

Sarah L. Kinnings¹, Nina Liu², Peter J. Tonge², Richard M. Jackson¹, Lei Xie^{3,4,*}, and Philip E. Bourne^{3,*}

¹Institute of Molecular and Cellular Biology and Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, United Kingdom

²Institute of Chemical Biology & Drug Discovery, Department of Chemistry, Stony Brook University, Stony Brook, NY 11794, USA

³Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA

⁴Department of Computer Science, Hunter College, The City University of New York, New York, NY 10065, USA

Abstract

Docking scoring functions are notoriously weak predictors of binding affinity. They typically assign a common set of weights to the individual energy terms that contribute to the overall energy score, however, these weights should be gene family-dependent. In addition, they incorrectly assume that individual interactions contribute towards the total binding affinity in an additive manner. In reality, noncovalent interactions often depend on one another in a nonlinear manner. In this paper we show how the use of support vector machines (SVMs), trained by associating sets of individual energy terms retrieved from molecular docking with the known binding affinity of each compound from high-throughput screening experiments, can be used to improve the correlation between known binding affinities and those predicted by the docking program eHiTS. We construct two prediction models; a regression model trained using IC₅₀ values from BindingDB, and a classification model trained using active and decoy compounds from the Directory of Useful Decoys (DUD). Moreover, to address the issue of overrepresentation of negative data in high-throughput screening data sets, we have designed a multiple-planar SVM training procedure for the classification model. The increased performance that both SVMs give when compared with the original eHiTS scoring function highlights the potential for using nonlinear methods when deriving overall energy scores from their individual components. We apply the above methodology to train a new scoring function for direct inhibitors of *M.tuberculosis* (*M.tb*) InhA. By combining ligand binding site comparison with the new scoring function, we propose that phosphodiesterase inhibitors can potentially be repurposed to target *M.tb* InhA. Our methodology may be applied to other gene families for which target structures and activity data are available, as demonstrated in the work presented here.

Introduction

Molecular docking aims to evaluate the feasible binding geometries of a putative ligand with a target of known 3D structure. Typically, docking algorithms consist of both a search algorithm for the exploration of different ligand (and sometimes protein) conformations, and

*To whom correspondence should be addressed: lxie@ucsd.edu or pbourne@ucsd.edu.

a scoring function for the calculation of ligand binding affinities. Ideally, the scoring function should be able to identify a solution with the correct ligand binding mode from alternative solutions, and ultimately be able to rank a set of ligands according to experimental binding affinity. In principle, the binding affinity should be calculated based on the first principle of thermodynamics. The most powerful approach is the absolute binding free energy (ABFE) approach¹⁻⁶, which uses extensive conformational sampling from molecular dynamics simulation, fully detailed atomic force fields, and a separate simulation of the solvation of the ligand, protein and associated complex. However, ABFE is too computationally expensive to be applied to screen millions of compounds. Furthermore, in spite of its cost, the prediction from ABFE is not always accurate⁷. Tremendous efforts have been made to develop physical-based or knowledge-based docking scoring functions to efficiently predict binding affinity. However, docking scoring functions remain notoriously weak predictors of binding affinity. Indeed, following an evaluation of 10 docking programs and 37 scoring functions, Warren et al.⁸ concluded that scoring functions are in need of significant improvements for predicting binding affinity. The principal reason for failure is the inability of the scoring function to reliably rank optimal 'native-like' ligand conformations above 'non-native' orientations⁹. Thus, although in most cases the correct binding mode can be retrieved during the conformational search, assigning the lowest energy score to the correct binding pose has proved to be more challenging. This inevitably leads to poor correlation with experimentally determined binding affinities. In general, the prediction of binding affinity is a challenging task since it is not only the result of collective weak noncovalent interactions, but it also includes the ability of the ligand to access the binding site, the desolvation free energy of the ligand and the binding site, and entropy and enthalpy changes in the ligand, protein, and solvent¹⁰. A realistic goal for docking scoring functions may be to discriminate active and inactive compounds and to rapidly filter out likely inactives in high-throughput screening campaigns.

Almost all existing docking scoring functions, including physical-based force fields, involve the fitting of data from experiments and calculations based on quantum mechanics. Docking scoring functions typically assign a common set of weights to the individual energy terms that contribute to the overall energy score, however, the weights assigned to the individual energy terms are, in reality, gene family-dependent. In order to achieve more accurate results, the docking score must be trained to derive a unique set of weights for each individual family. Furthermore, many docking algorithms incorrectly assume that individual interactions contribute towards the total binding affinity in an additive manner, and therefore derive their predictions of binding affinity from a linear combination of individual energy terms. However, this method is not theoretically sound, since it fails to consider the cooperative effects of noncovalent interactions. Such cooperativity has only been considered recently. These studies all highlight the need for docking scoring functions that are based on nonlinear models¹⁰⁻¹³. Thus, the development of target-dependent docking scoring functions based on nonlinear models would significantly enhance their predictive ability, facilitating the rational design of ligands that bind to specific sites on target proteins¹¹.

The ability of machine learning methods to produce effective scoring functions has been demonstrated in a number of recent studies. For instance, both Deng *et al.*¹⁴ and Ballester and Mitchell¹⁵ characterized protein-ligand complexes using feature vectors comprising the number of occurrences of specific protein-ligand atom type pairs interacting within predefined distance thresholds. While Deng *et al.* trained Kernel a PLS (partial least squares) to predict protein-ligand binding affinity, Ballester and Mitchell trained a random forest classifier for the same purpose. The algorithm by Ballester and Mitchell ('RF-Score') was shown to outperform 16 existing scoring functions when using a benchmark consisting of 195 diverse protein-ligand complexes. In 2005, Springer *et al.*¹⁶ presented PostDOCK, a post-processing filter to distinguish true binding protein-ligand complexes from docking

artefacts generated by the popular docking program DOCK 4.0.1¹⁷. PostDOCK uses biochemical descriptors to characterize the protein-ligand interaction, including nonbonded van der Waals and electrostatic interaction force field terms from the DOCK score, terms for hydrogen bonding, metal bonding, lipophilicity, and a series of buried solvent accessible surface area descriptors. The authors used a random forest classifier to separate the binding and the nonbinding ligands from a test set of 44 structurally diverse protein targets, and showed that PostDOCK was able to outperform both the DOCK and ChemScore¹⁸ scoring functions. More recently, Sato *et al.*¹⁹ developed Pharm-IF, a novel atom-based, pharmacophore-based interaction fingerprint. The fingerprint is calculated from the distances of pairs of ligand pharmacophore features that interact with protein atoms. Hydrogen bonds, ionic interactions and hydrophobic interactions are considered. Four machine learning algorithms (naïve Bayesian classifier, random forest, support vector machine and artificial neural network) were used to build prediction models based on Pharm-IF, and each was shown to outperform the GLIDE²⁰ docking algorithm across all five targets investigated. Durrant and McCammon²¹ also recently demonstrated that neural networks can be used to successfully characterize the binding affinities of protein-ligand complexes. Their neural networks, which were trained by characterizing 4,141 protein-ligand complexes across 194 dimensions, were used to rescore the docked poses of a selected set of ligands. The authors demonstrated that their networks were able to distinguish between well-docked and poorly docked ligands, and even between true ligands and decoy compounds when both were well-docked.

Classification support vector machines (SVMs) separate positive and negative data in multidimensional space by drawing an optimal hyperplane with a maximum distance between the two classes. As such, SVMs can potentially be applied in developing nonlinear models of docking scoring functions using high-throughput screening (HTS) data. Unfortunately, data imbalance, which thwarts SVMs, is a common feature amongst a wide range of experimental data, including the data here. It occurs when one of the classes in the dataset, commonly the class of interest, is represented by a significantly smaller number of samples than the other class(es). For instance, of the tens of thousands of compounds typically investigated in HTS, it is likely that only a small fraction will show activity against the target of interest. As a result, HTS data are commonly imbalanced, with the vast majority of compounds belonging to the negative class. The use of such imbalanced data in the training of classification SVMs tends to push the optimal hyperplane towards the positive class, resulting in a tendency of the model to classify observations into the negative class, and therefore have high precision but low recall²².

Sampling strategies, such as oversampling the minority class, or undersampling the majority class are popular choices when addressing the issue of data imbalance²³. In a recent study, Tang *et al.*²⁴ proposed a granular sampling strategy for data rebalancing, which undersamples the majority class, while minimizing information loss. Their method treats all positive samples as important due to their rarity, and discards uninformative (i.e., non-SV) negative samples. Indeed, since only SVs are important for SVM model classification, the removal of non-SV samples has no significant effect on model performance. In this way, the granular sampling method identifies samples that are adjacent to the border between the two classes, and uses them to draw the separating hyperplane. Those samples that are far away from the hyperplane are discarded during model building. Li *et al.*²² recently applied this granular sampling method to facilitate the analysis of highly imbalanced luciferase inhibition bioassay data. While the data imbalance associated with HTS data has been noted in a number of other studies²⁵⁻²⁷, to our knowledge, there have been no methods reported to tackle this problem effectively.

To address the issue of data imbalance in SVM training during docking, we have developed a multiple planar SVM scheme to balance precision and recall. The multiple planar SVM randomly partitions the overrepresented negative cases into n subsets. n SVM models are then trained using each of the partitioned negative subsets, plus a common set of positive cases. The prediction is based on the combined score of the n SVM models (Figure 1). Compared to the granular sampling method, the multiple planar method is more computationally efficient. Indeed, the main drawback of the granular sampling method is that it is not possible to extract all informative negative samples using a single SVM. For instance, in the aforementioned work of Li et al.²², 67 sampling rounds, each involving the modelling of a new SVM, were required before the results converged and the sampling process was considered complete. While the multiple planar method also requires the training of multiple SVM models, each of these models is a small fraction of the size of the overall dataset.

In this paper we show how the use of a SVM, trained by associating sets of individual energy terms retrieved from molecular docking with the known binding affinity of each compound in the training set, can be used to improve the correlation between known binding affinities and those predicted by a popular docking program for a series of known inhibitors. The docking program eHiTS²⁸ was selected for this study due to its ability to output a large number of individual energy terms contributing to the overall energy score. Although we only used eHiTS in this study, our methodology can be applied to any other docking software provided that it outputs individual energy terms. More generally, it is applicable to all docking algorithms if 3D structural fingerprints are used as feature vectors.

We have applied this methodology to the development of an enoyl acyl carrier protein reductase (InhA)-specific docking scoring model. InhA is the target of the first-line anti-tubercular drug isoniazid. The therapeutic effect of isoniazid relies on it forming a covalent conjugate with an NAD co-factor. Since the extensive use of isoniazid has prompted the emergence of drug-resistant *M.tb* strains, the development of direct inhibitors of InhA is highly desirable²⁹. We are particularly interested in repurposing existing pharmaceuticals to directly inhibit InhA. Indeed, in a recent study, we have shown that the safe pharmaceutical Comtan, a catechol-*O*-methyltransferase (COMT) inhibitor, which is used in the treatment of Parkinson's disease, is able to directly inhibit InhA³⁰. In this report, based on the SVM model and ligand binding site similarity, we propose that phosphodiesterase inhibitors could potentially be direct inhibitors of InhA.

Results

SVM regression model

Firstly, we developed a regression support vector machine (SVM), which was calibrated using IC₅₀ data from 80 InhA inhibitors in BindingDB 31 (see Methods section). The individual energy terms from the eHiTS docking software²⁸ were used as features to train the model. The initial correlation between the eHiTS-predicted binding affinities (which were derived from a linear combination of energy terms) and the experimentally determined IC₅₀ values of 80 InhA inhibitors from BindingDB was very poor (Spearman's rank correlation coefficient $\rho = 0.117$).

The eHiTS output provides a total of 20 different energy terms that contribute to the overall energy score (*H_bond*, *Lipophil*, *pi_stack*, *solvent*, *steric*, *strain*, *Coulomb*, *entropy*, *other*, *Rcharge*, *Rshape*, *RlogD*, *Rcover*, *Lcharge*, *LlogD*, *Lcover*, *metal*, *lig_int*, *family* and *depth*). In order to determine the optimal combination of energy terms (i.e., features) for the regression model, different combinations were tested. All 20 energy terms were extracted from the output resulting from the docking of 80 different InhA inhibitors into InhA. As

stated above, these inhibitors were selected because the IC₅₀ values of their binding to InhA had been experimentally determined, and could therefore be associated with their respective feature vectors during SVM training.

The terms *metal* and *lig_int* were discarded since they were of zero value for all inhibitors. The terms *family* and *depth* were also discarded since they relate to the protein structure only, and since this was constant, they would simply add noise to the data. Of the remaining 16 energy terms, nine were regarded as compulsory. These included terms that are directly involved in the molecular interaction, such as hydrogen bonding (*H_bond*), hydrophobic (*Lipophil*), pi-pi stacking (*pi_stack*), solvation (*solvent*), van der Waals (*steric* and *strain*), electrostatics (*Coulomb*), and entropy (*entropy*). The seven remaining terms were regarded as optional (*Rcharge*, *Rshape*, *RlogD*, *Rcover*, *Lcharge*, *LlogD* and *Lcover*). Each combination tested consisted of the nine compulsory energy terms plus a different combination of the seven optional energy terms, resulting in a total of 128 different combinations.

Five-fold cross-validation was then used to compare the relative accuracies of the 128 different combinations of features. For each combination, there were five iterations where, each time, four subsets of equal size were used as training data, and the remaining subset was used as test data. Spearman's rank correlation coefficient (ρ) was calculated for each iteration, and the mean correlation coefficient across all five iterations was assigned to each combination. The combination of energy terms that resulted in the highest mean correlation coefficient was chosen for the model. The best combination of features consisted of the nine compulsory energy terms, plus *Lcover* and *Rcharge*.

The best combination of features resulted in a significant improvement in the correlation coefficient ($\rho = 0.607$), when compared with that achieved by the eHiTS scoring function ($\rho = 0.117$). To test if the model was over-fitted, y-randomization³² was carried out. The highest correlation coefficient obtained when carrying out 100 y-randomization trials (see Methods section) was 0.384 (mean = 0.079), which is significantly lower than $\rho = 0.607$ (Figure 2). This suggests that the regression model would be able to more accurately predict the binding affinities of a series of compounds with InhA than the eHiTS scoring function would alone. Indeed, unlike the eHiTS scoring function, the SVM does not assume the addition of the individual energy terms. Instead, for each molecule, it uses an experimentally determined IC₅₀ value to weight these energy terms according to their likely contribution to the overall binding affinity. In this way, the cooperative effects of noncovalent interactions may be taken into account.

A multiple-planar SVM classification model

Huang *et al.* recently constructed the Directory of Useful Decoys (DUD)³³, in which the decoys were specifically selected to be physicochemically similar yet topologically dissimilar to the active ligands, thereby providing a more realistic scenario for the validation of ligand-based virtual screening. The DUD provides a useful resource for training SVM classification models, because the feature vector of each molecule can be labelled as either an active (+1) or a decoy (-1). However, since it contains actives and decoys in a ratio of approximately 1:36, there is a strong bias towards negative examples in the training data. In order to overcome this problem, a multiple-planar classification model was developed. After randomly dividing the actives and decoys into five subsets for the five-fold cross-validation, each of the five subsets of decoys was randomly partitioned into a further n subsets. In order to determine the optimal value of n , five-fold cross-validation was carried out for various values of n from five to 36. Five-fold cross-validation was carried out in the same way as for the regression model, except that for each iteration, n models were trained using a different subset of decoys.

In a classification SVM, two classes are separated by the value zero, and a positive score indicates a positive class, whereas a negative score indicates a negative class. The absolute value of the score is comparable to the distance from the hyperplane. Since each of the n models predicted a score for each of the molecules in the test set, a weighted voting method was used where the sum of the n scores was assigned to each molecule. If the sum of the n scores was greater than zero, the molecule was predicted to be an active, whereas if the sum of the n scores was less than zero, the molecule was predicted to be a decoy. The F -score was then calculated (see the Methods section) for each of the five iterations, and the mean F -score was assigned to that particular value of n . Figure 3 shows that the mean F -score increases rapidly as n is increased to 15, where it hits a peak of 0.36, before it starts decreasing steadily until n reaches 36. Therefore, the number of partitions in the decoy set was set to 15 for the classification model.

The performance of the eHiTS scoring function and that of the classification SVM at discriminating actives from decoys is shown in the ROC curve in Figure 4. Although eHiTS provides some initial early enrichment with respect to random, with an enrichment factor of 4.9 at the 1% level, it actually performs worse than random throughout the entire remainder of the dataset. The classification model, on the other hand, performs well across the entire dataset, with enrichment factors of 6.0, 8.8, and 6.5 at the 2%, 5% and 10% levels, respectively. It is true that only 0.01-1% compounds are selected from a huge compound database in the drug discovery process. However, in our test case, only hundreds of compounds are screened. Therefore, 1% of the test data set is not statistically significant, as suggested by Sato *et al.*¹⁹. Furthermore, in order to directly compare the performance of the classification SVM with that of the regression SVM, the regression SVM was used to rank the DUD InhA data set according to predicted $\log IC_{50}$ values. The resulting ROC curve (Figure 4) shows that while the regression model significantly outperforms the eHiTS scoring function, it is unable to match the performance of the classification model. It is likely that the higher accuracy of the classification model, when compared to that of the regression model, is due to the inclusion of negative data in the training set. Indeed, since the regression model was developed purely on the basis of positive data, the energy terms for improper interactions would have received little weight during parameter estimation. Therefore, by incorporating negative training data, the classification model allows for the rigorous estimation of all scoring function parameters³⁴.

Relative contributions of energy terms

A feature selection procedure was carried out (see Methods section) in order to compare the relative contributions of the nine compulsory energy terms to the predictions made by both the regression and classification models. The resulting Spearman's rank correlation coefficients (ρ) and F -scores are shown in Table 1 for the regression and classification models, respectively. There is considerable overlap between the relative contributions of the nine different energy terms for the regression and classification models. The hydrophobic, pi-pi stack, solvation and van der Waals steric energy terms were determined to make strong contributions towards the overall binding affinity prediction in both models. Indeed, their exclusion from the optimized models resulted in a significant decrease in performance when compared with that of the optimized models (labelled 'none removed' in Table 1). Conversely, the removal of the electrostatics, entropy, other and van der Waals strain energy terms from the optimized models had a much smaller effect on performance, therefore suggesting that these energy terms made less important contributions to the predictions made by both models.

Interestingly, removal of the hydrogen bonding energy term had a much larger effect on the performance of the regression model than on that of the classification model. Indeed, its removal resulted in a significant decrease from $\rho = 0.637$ to $\rho = 0.110$ for the regression

model, yet a relatively smaller decrease from $F = 0.232$ to $F = 0.166$ for the classification model. Many existing direct InhA inhibitors form a hydrogen bond with the Tyr158 residue of InhA, and since the regression model was trained using only direct InhA inhibitors, this may explain the importance of the hydrogen bonding energy term to the regression model. In addition to direct InhA inhibitors, the classification model included a large number of decoys in the training data. Therefore, the energy terms contributing towards improper interactions would have received a greater weight in the classification model than in the regression model, thus explaining the reduced importance of the hydrogen bonding energy term in the classification model.

Overall, these results suggest that the eHiTS scoring function gave too little weighting to the important hydrophobic, pi-pi stack, solvation and van der Waals steric energy terms, yet too much weighting to the less important electrostatics, entropy, other and van der Waals strain energy terms. The results also suggest that the regression model gave too much weighting to the hydrogen bonding energy term, probably due to the fact that it was developed purely on the basis of positive data. This difference in the weighting of the hydrogen bonding energy term may account for the reduced performance of the regression model, relative to the classification model, as observed in Figure 4.

Potential inhibition of InhA by phosphodiesterase inhibitors

InhA is the primary molecular target of isoniazid, which has been used as a frontline anti-tubercular agent for the past 40 years^{35, 36}. As a prodrug, the activity of isoniazid is dependent on activation by KatG, a catalase/peroxidase enzyme. Unfortunately, it is this activation requirement that allowed *M.tb* to acquire resistance to the drug. Indeed, mutations in the katG gene account for around half of all isoniazid-resistant clinical isolates. Direct InhA inhibitors that avoid this activation requirement are not susceptible to this resistance mechanism. Triclosan and the diazaborines are well-known InhA inhibitors that do not require activation, but they are not suitable for human treatment due to their respective poor solubility and toxicity^{30, 37}. Therefore, novel, direct inhibitors of InhA are urgently required.

To identify potential lead compounds from existing drugs that are able to directly inhibit InhA, we used our binding site similarity comparison software SMAP³⁸⁻⁴⁰ to search InhA against 962 protein structures that are co-crystallized with 274 FDA approved drugs⁴¹ in the RCSB Protein Data Bank (PDB)⁴². With SMAP P -values as low as $4.36e-8$ (Table 2), some of the strongest connections to InhA were with the phosphodiesterase type 5 (PDE5) inhibitors, which are used in the treatment of erectile dysfunction. Although only PDE5 is co-crystallized with approved drugs in the PDB, a number of inhibitors have been developed to target other members in the PDE gene family. Therefore, 303 PDE inhibitors from BindingDB were docked into the InhA binding site (including the NAD co-factor as a part of the receptor) to identify the most potent inhibitors.

Three drugs vardenafil, tadalafil and sildenafil all showed highly favourable eHiTS energy scores of as low as -11.74 pKi (Table 2), which ranks them in the top three among 303 PDE inhibitors. When using the SVM regression model they only ranked at 83/303, 258/303 and 299/303, respectively. An enzyme kinetics assay was carried out in order to test the activity of vardenafil against InhA. No inhibition was observed up to $200 \mu\text{M}$. This implies that the SVM model is able to improve the ranking accuracy of the original eHiTS scoring function. However, the fact that vardenafil has a predicted IC_{50} of $1,545 \text{ nM}$ ($1.545 \mu\text{M}$) when it has been experimentally determined to be a non-binder, suggests that the absolute IC_{50} values predicted by the SVM may not be entirely accurate. Even so, the best predicted IC_{50} values among the 303 PDE inhibitors are more than two-fold lower than that of vardenafil, and so they may still be potential inhibitors of InhA (Table 3). Interestingly, all top 10 ranked compounds are PDE4 inhibitors. When the 303 PDE inhibitors from BindingDB were

ranked using the classification model (Table 4), there was a considerable overlap between the top ranked compounds and those with the best predicted IC_{50} values from the regression model. In particular, there are 6 PDE4 inhibitors that rank within the top 10 compounds when using either model; L-869299, L-869298, triazolothiadiazine 8, phthalazinone 22, V11294 analog 5u and V11294 analog 5t. Experimental testing is required to verify whether or not these PDE4 inhibitors are active against InhA. Note that with an original eHiTS energy score of -9.83 pKi, triazolothiadiazine 8 would not have been ranked particularly highly when using the eHiTS scoring function.

The binding pose of L-869298 within the InhA substrate binding site, as predicted by eHiTS, is shown in Figure 5. It overlaps well with the crystallographically-determined binding pose of the direct, high affinity InhA inhibitor (3S)-1-cyclohexyl-N-(3,5-dichlorophenyl)-5-oxopyrrolidine-3-carboxamide (PDB code: 2h7m, PDB ligand code: 641, IC_{50} : 390nM). Note that information about the cocrystallized conformations of known inhibitors was not used in any part of the docking process. Further investigation of the predicted binding pose of L-869298 revealed two potential hydrogen bonds, one between a fluorine atom (F13) of L-869298 and the amide nitrogen of a leucine residue (Leu197) in InhA, and the other between the same fluorine atom and the amide nitrogen of an alanine residue (Ala198) in InhA. Hydrophobic interactions were also observed between L-869298 and the following InhA residues; Gly96, Met103, Tyr158, Met161, Thr196, Leu197, Ala198, and Met199.

Following the PDE5 inhibitors, the next most significant connection to InhA predicted by SMAP was with the sex hormone estradiol, which binds to estrogen receptors (ERs) α and β . Furthermore, at a SMAP *P*-value threshold of $1.0e-5$, InhA was also connected to the selective ER modulator raloxifene. These findings suggest that compounds that bind to ERs may also have the potential to directly inhibit InhA. There are a total of 223 compounds in BindingDB that have been confirmed experimentally to bind with any of ER α , ER β , estrogen-related receptor (ERR) α or ERR γ . Both the regression and the classification model of InhA were used to rank these compounds according to the probability of their activity against InhA (see the Supporting Information, Tables S1 and S2). As with the PDE inhibitors, there was a strong agreement between the two models, therefore providing further support for the results. Interestingly, the most highly ranked compounds were analogs of 4-hydroxytamoxifen, the active metabolite of tamoxifen, which were predicted by the regression model to fall within the nanomolar range. Again, these compounds would be ideal candidates for experimental testing against InhA.

The multiple-planar SVM classification approach is applicable to a diverse set of targets

Since the DUD provides sets of actives and decoys for a number of different targets, multiple-planar classification models were developed for 12 additional targets, in order to test the robustness of the methodology. For target selection criteria, please refer to the Methods section. Figure 6 shows ROC curves comparing the performance of the different classification models (purple lines) with that of the eHiTS scoring function (green lines) for the 12 selected targets. The classification models can be seen to outperform the eHiTS scoring function across every single target. In all cases except for HIV reverse transcriptase (hivrt), the classification models provide significant early enrichment. For hivrt, the eHiTS scoring function performs worse than random (black line), while the classification model barely performs above random across the entire dataset. The eHiTS scoring function, on the other hand, can be seen to perform close to, or worse than, random across the majority of targets. The eHiTS scoring function gives its best performance for cyclooxygenase-2 (cox2), but even then, it is significantly outperformed by the classification model, which also gives its best performance for this target. It is interesting to note that hivrt is the target with the least number of actives (15) in the dataset, while cox2 is the target with the greatest number

of actives in the dataset (43). The classification model also performs extremely well for epidermal growth factor receptor (egfr), which is the target with the second greatest number of actives (40), however, the eHiTS scoring function performs worse than random for this target. This may be because larger training sets cover more extensive spaces of both positive and negative examples.

Discussion

We have trained two different SVMs for an important anti-tubercular target, InhA, by associating sets of eHiTS-predicted individual energy terms with the known binding affinity of each compound in the training set. In this way, the docking scoring function was trained to derive a unique set of weights specific to InhA. Since the methodology does not require cocrystallized structural information, it can take full advantage of the vast amounts of data generated by HTS experiments. The abundance of data from HTS means that it is now feasible to iteratively train a scoring function for individual gene families. In fact, we have extended the classification SVM to 12 different targets, and shown that it is able to outperform the eHiTS scoring function across every target. Although there are a limited number of targets for which there is sufficient experimental training data at present, many of these targets are important validated drug targets. Furthermore, the potential of our method can only improve as the availability of experimental data increases in the future.

Unfortunately, a major hurdle in the use of HTS data to train classification SVMs is the low ratio of active to inactive compounds. Such imbalanced data results in a strong bias towards the prediction of negative cases, therefore resulting in a model with high precision but low recall. We have presented a novel and computationally efficient multiple-planar approach to overcome this previously acknowledged but frequently neglected issue²². Our multiple-planar approach may be applied to train classification SVMs using any set of imbalanced data. Imbalanced data is not only an issue for bioinformatics, but also for applications as wide-ranging as e-business, information security and national security²³.

Furthermore, by using SVMs to model the individual energy terms in a nonlinear fashion, it may be possible to implicitly simulate their cooperative behaviour. Indeed, noncovalent interactions are frequently either positively or negatively cooperative in nature: They exhibit positive cooperativity when the binding energy derived from them acting together is greater than would be derived from the sum of them acting separately, and exhibit negative cooperativity when the opposite situation is true¹³. Unfortunately, docking scoring functions, which are primarily designed for high throughput analyses, tend to ignore this fundamental concept of cooperativity in favour of the simpler and more convenient additivity approach¹¹. A number of recent studies have highlighted the need for docking scoring functions that incorporate descriptors that model cooperative effects¹⁰⁻¹³. In particular, Muley et al.¹⁰ suggested that the potential error in binding affinity predictions resulting from ignoring the cooperativity between a hydrogen bond and a hydrophobic side chain of 100 Å² contact surface could be in the region of one order of magnitude. It is noted that cooperativity in protein-ligand interactions appears at the atomic level. Unfortunately, the use of energy terms derived from docking software does not fully capture the atomic details of protein-ligand interactions. Therefore, the incorporation of atomic interaction features into the computational framework presented in this paper may provide a better model of cooperativity in the future.

Although our method has only been applied to eHiTS thus far, the promising results obtained from this study indicate that other docking scoring functions or interaction fingerprints would benefit from such treatment. Indeed, the fact that scoring functions are notoriously weak predictors of binding affinity can be partly attributed to their neglect to

derive a unique set of weights specific to each gene family, and also to the additive approach that they employ to derive overall energy scores from their individual contributions. We hope that our endeavours in this paper will promote a much needed review of the theory behind current docking scoring functions, therefore paving the way for their long overdue improvement.

Conclusions

We were able to increase the accuracy of virtual screening of direct InhA inhibitors by using SVMs, in which the features of the scoring matrix were energy terms retrieved from molecular docking using eHiTS. We constructed two prediction models for InhA; a regression model, which was trained using InhA experimental IC₅₀ values from BindingDB, and a classification model, which was trained using the InhA DUD dataset. The two models gave a good consensus. Our results based on y-randomization indicate that the improved ranking performance is not due to over-fitting, and that our model may capture new information that is missed by the eHiTS docking scoring function, which is considered to be the state-of-the-art, but incorrectly assumes that individual interactions contribute towards the total binding affinity in an additive manner. The purpose of our new model is to capture the non-linear cooperative features of noncovalent interactions. Although there is still room for improvement as far as our methodology is concerned, i.e., the fact that the features in our model do not explicitly capture the atomic details of protein-ligand interactions, we hope that our work may inspire future endeavors in developing new scoring functions that correctly model the cooperative effects of protein-ligand interactions. Furthermore, our multiple-planar SVM training procedure for the classification model successfully addressed the issue of overrepresentation of negative data in the high-throughput screening data set. Using our SVMs we were able to produce a prioritized list of PDE inhibitors for experimental screening against InhA. In addition, our SVMs were used to prioritize a set of compounds known to bind estrogen receptors and estrogen-related receptors. In principle, our methodology could be directly applied to other targets of interest, as demonstrated in the work presented here, and also to other docking programs.

Methods

Molecular docking of the PDE5 inhibitors into InhA

Since the SMAP software had aligned each drug binding site with the InhA binding site, the aligned coordinates of each drug molecule were used to define the search space for docking each drug into the *M.tb* protein (i.e., the aligned drug molecule was used as the *clip* file with a default search space of 10 Å³). As recommended by the manual, the eHiTS accuracy level was increased to 6 (default = 3), in order to improve the accuracy of the predictions. Following all docking, the binding pose with the highest estimated binding affinity was selected for further investigation. Note that the NAD cofactor was added as the last residue in the InhA structure prior to docking.

Normalization of PDE5 inhibitor docking scores with InhA

In order to normalize a docking score, 500 Lipinski-compliant⁴³ molecules with the same number of carbon atoms, the same number of nitrogen atoms and the same number of oxygen atoms as the drug of interest were randomly selected from subset 13 of the ZINC database⁴⁴, and docked into the target of interest using the same protocol. The means and standard deviations of the three sets of 500 resulting energy scores were then used to derive a Z-score for the energy score of each of the three drugs according to the following formula:

$$Z\text{-score} = (\text{energy score} - \text{mean of 500 energy scores}) / \text{std. dev. of 500 energy scores}$$

The *P*-value was then derived from the *Z*-score using a normal distribution with a mean and standard deviation of the 500 energy scores.

Regression SVM

BindingDB³¹ (<http://www.bindingdb.org/>) provides a rich repository of experimentally-measured data for protein-ligand interactions, including IC₅₀ values for 98 different direct InhA inhibitors (October 8, 2009). The IC₅₀ values of 21 of these inhibitors are not exact (i.e., >100,000 nM) and so were excluded. In addition, IC₅₀ data for four other direct InhA inhibitors⁴⁵ that are present in the RCSB PDB but not BindingDB were included, resulting in IC₅₀ values for 81 different direct InhA inhibitors. For the 14 of these inhibitors that were cocrystallized with InhA in the RCSB PDB, FATCAT⁴⁶ (<http://fatcat.burnham.org/>) was used to structurally align the ligand-bound chains of the InhA structures onto 1BVR, a structure of InhA bound with a fatty acid substrate (PDB ligand code: THT). The aligned inhibitor molecules were then extracted and docked into 1BVR using eHiTS²⁸, using the THT substrate to define the docking search space. For each inhibitor, the energy terms of the binding pose with the smallest RMSD to the cocrystallized ligand were retained. The 67 remaining inhibitor molecules were downloaded from BindingDB in sdf format, and also docked into 1BVR using the fatty acid substrate to define the docking search space. For each of these inhibitors, the energy terms of the binding pose with the greatest binding affinity (i.e., lowest energy score) were retained. Note: One inhibitor, 14PP, failed to dock. Spearman's rank correlation coefficient was then calculated between the actual IC₅₀s and the eHiTS predicted energy scores of the 80 InhA inhibitors.

SVM-light⁴⁷ was used to train a regression model where each of the 80 InhA inhibitors was labelled with the logarithm of its IC₅₀ value and was assigned a feature vector comprising the eHiTS energy terms. The optimal combination of features was determined using five-fold cross-validation (see Results section for further details). In order to determine the optimal kernel function for this combination of features, three other kernel functions (*polynomial*, *radial basis function* and *sigmoid tanh*), besides the default *linear* kernel, were tested and mean correlation coefficients were calculated. However, the *linear* kernel gave the best results and so was chosen for the model. All five subsets were then merged to create the regression model using the chosen combination of energy terms and the *linear* kernel function.

All 303 PDE inhibitors present in BindingDB (on October 16, 2009) were downloaded in sdf format, and docked into InhA using eHiTS. Again, the 1BVR structure and the THT substrate were used to define the docking search space. Note that one inhibitor, CI-930, failed to dock. The chosen combination of energy terms was extracted from the eHiTS output for each inhibitor. SVM-light was then used to predict the log IC₅₀ value of each PDE inhibitor with InhA, based on the regression model. The 303 PDE inhibitors were then ranked based on their predicted log IC₅₀ values when binding to InhA. This process was repeated using 223 molecules from BindingDB that bind estrogen receptors and estrogen-related receptors.

Classification SVM

All 85 active molecules and 3,035 decoy molecules for InhA were downloaded from the DUD³³ (<http://dud.docking.org/>). These molecules were docked into the 1BVR structure of InhA, using the THT substrate as a clip file. Note that 11 decoy molecules failed to dock,

leaving a total of 3,024 decoys. SVM-light was used to train a classification model where each molecule was labelled as either an active (+1) or a decoy (-1) and was assigned a feature vector comprised of eHiTS energy terms. The classification model was trained using the combination of features and kernel function that were selected previously for the regression model. The optimal number of partitions for the decoys was determined using five-fold cross-validation (see the Results section for further details).

In order to compare the performance of eHiTS (using the default energy scores) with that of the classification model, receiver operator characteristic (ROC) curves were plotted. ROC curve analysis is considered to be the best approach for characterizing the performance of virtual screening methods 48. However, since the DUD is subject to analog bias amongst the actives, this can lead to falsely high enrichment rates. Indeed, in its raw state, there are many targets for which nearly all actives are trivial analogs of a central structure 49. In a recent study, Good and Oprea 50 filtered the DUD actives for lead-likeness (molecular weight (MW) < 450 and $\log P < 4.5$) before clustering them according to chemotype using a reduced graph representation. In this way, they not only removed large molecules with unsuitable physicochemical properties, but they also addressed the problem of analog bias. For InhA, this procedure resulted in the removal of 28 actives, leaving 57 actives, which were then clustered into 23 clusters (<http://dud.docking.org/clusters/>, accessed November 11, 2009). In order to prevent analog bias in the ROC curves, the filtered and clustered set of InhA actives was used in combination with an arithmetic weighting procedure where each active molecule was given a weighting that was equal to the inverse of the size of the cluster to which it belonged^{51, 52}. For instance, if five active molecules with similar chemotypes had been assigned to the same cluster, each active within that cluster would be given a weighting of one-fifth. Exactly the same procedure was used to evaluate the ability of the regression model to correctly rank the actives and decoys in the InhA DUD dataset.

In the same way as before, SVM-light was used to predict whether each of the 303 PDE inhibitors would be able to bind InhA, based on the classification model. The 303 PDE inhibitors were then ranked based on the absolute values of their scores. This process was also repeated using 223 molecules from BindingDB that bind estrogen receptors and estrogen-related receptors.

y-Randomization

y-Randomization is a popular model validation tool whereby the performance of the original model is compared to that of models built from randomly shuffled responses, based on the original descriptor pool and the original model building procedure. In this way, it is possible to protect against the risk of chance correlation³². In order to implement a y-randomization procedure, the y values (i.e., $\log IC_{50}$ values) were randomly re-ordered for all 80 inhibitors in the training set for the regression SVM. The training procedure was then carried out in exactly the same way as before, i.e., five-fold cross-validation for 128 combinations to determine the optimal combination of features. When calculating the correlation coefficient for each of the five subsets in each of the 128 combinations, the ranks of the new y-randomized $\log IC_{50}$ values were compared with the ranks of the predicted $\log IC_{50}$ values from the y-randomized model. This process was repeated 100 times, and each time the highest mean correlation coefficient from five-fold cross-validation, achieved across the 128 combinations of features, was reported.

Measurement of accuracy using F-scores

$F\text{-score} = (2 \times \text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$ Where; precision equals the number of correct positive results divided by the total number of predicted positive results, and recall

equals the number of correct positive results divided by the total number of results that should have been predicted as positive.

Feature selection procedure

A feature selection procedure was carried out in order to compare the relative contributions of the nine compulsory energy terms to the predictions made by both the regression and classification models. In each case, a model was created using the optimal combination of features from the entire set of training data, and this model was then used to make predictions about the entire set of training data. Note that for the classification SVM, the multiple-planar method was employed with the optimal number of partitions. 10 different models were created for each of the regression and classification SVMs; one where all 11 predetermined features were included in the model, and nine where a different compulsory energy term had been removed from the model. All 10 models were used to make predictions about the same set of training data, and Spearman's rank correlation coefficients and *F*-scores were calculated for the regression and classification SVMs, respectively.

Development of SVM classification models for 12 additional targets

The DUD contains actives and decoys for 40 different targets. As mentioned previously, Good and Oprea⁵⁰ filtered and clustered each set of actives in order to remove large molecules with unsuitable physicochemical properties, and also to address the problem of analog bias. To restrict our analysis to targets with a sufficient number of chemotypes, only those targets possessing at least 15 different clusters of actives were selected (as in the work by Cheeseright *et al.* 53 and Kinnings and Jackson 52). The sets of active and decoy molecules corresponding to each of the 12 selected target proteins (ace, ache, cdk2, cox2, egfr, fxa, hivrt, p38, pde5, pdgfrb, src and vegfr2) were downloaded directly from the DUD (<http://dud.docking.org/r2/>; DUD release 2, 22/10/2006) in mol2 file format. For each target, the protein structure used in the DUD benchmarking process³³ was selected, and the corresponding ligand was used to define the docking search space. All actives and decoys were docked into their relevant target structures, and the energy terms of the binding pose with the greatest binding affinity (i.e., lowest energy score) were retained for each molecule. For each target, the optimal combination of energy terms *and* the optimal number of partitions of decoys was determined for the classification model, using the methods described previously for the InhA classification model. Again, in order to prevent analog bias in the ROC curves, an arithmetic weighting procedure was used, where each active molecule was given a weighting that was equal to the inverse of the size of the cluster to which it belonged.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We appreciate the constructive suggestions of the anonymous reviewers and the editor in improving the manuscript. We are grateful for financial support from the National Institutes of Health grant GM078596. S.L. Kinnings was funded by a BBSRC studentship.

References

1. Luo H, Sharp K. On the Calculation of Absolute Macromolecular Binding Free Energies. *Proc. Natl Acad. Sci. U S A.* 2002; 99:10399–404. [PubMed: 12149474]
2. Boresch S, Tettinger F, Leitgeb M, Karplus M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B.* 2003; 107:9535–9551.

3. Woo HJ, Roux B. Calculation of Absolute Protein-Ligand Binding Free Energy from Computer Simulations. *Proc. Natl Acad. Sci. U S A.* 2005; 102:6825–30. [PubMed: 15867154]
4. Wang J, Deng Y, Roux B. Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophys. J.* 2006; 91:2798–814. [PubMed: 16844742]
5. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *J. Mol. Biol.* 2007; 371:1118–34. [PubMed: 17599350]
6. Roderger T, Howell PL, Pomes R. Calculation of Absolute Protein-Ligand Binding Free Energy Using Distributed Replica Sampling. *J. Chem. Phys.* 2008; 129:155102. [PubMed: 19045232]
7. Mobley DL, Dill KA. Binding of Small-Molecule Ligands to Proteins: “What You See” Is Not Always “What You Get”. *Structure.* 2009; 17:489–98. [PubMed: 19368882]
8. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* 2006; 49:5912–31. [PubMed: 17004707]
9. Korb O, Stutzle T, Exner TE. Empirical Scoring Functions for Advanced Protein-Ligand Docking with Plants. *J. Chem. Inf. Model.* 2009; 49:84–96. [PubMed: 19125657]
10. Muley L, Baum B, Smolinski M, Freindorf M, Heine A, Klebe G, Hangauer DG. Enhancement of Hydrophobic Interactions and Hydrogen Bond Strength by Cooperativity: Synthesis, Modeling, and Molecular Dynamics Simulations of a Congeneric Series of Thrombin Inhibitors. *J. Med. Chem.* 2010; 53:2126–35. [PubMed: 20148533]
11. Baum B, Muley L, Smolinski M, Heine A, Hangauer D, Klebe G. Non-Additivity of Functional Group Contributions in Protein-Ligand Binding: A Comprehensive Study by Crystallography and Isothermal Titration Calorimetry. *J. Mol. Biol.* 2010; 397:1042–54. [PubMed: 20156458]
12. Lafleur K, Huang D, Zhou T, Caflisch A, Nevado C. Structure-Based Optimization of Potent and Selective Inhibitors of the Tyrosine Kinase Erythropoietin Producing Human Hepatocellular Carcinoma Receptor B4 (Ephb4). *J. Med. Chem.* 2009; 52:6433–46. [PubMed: 19788238]
13. Williams DH, Stephens E, O'Brien DP, Zhou M. Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angew. Chem. Int. Ed. Engl.* 2004; 43:6596–616. [PubMed: 15593167]
14. Deng W, Breneman C, Embrechts MJ. Predicting Protein-Ligand Binding Affinities Using Novel Geometrical Descriptors and Machine-Learning Methods. *J. Chem. Inf. Comput. Sci.* 2004; 44:699–703. [PubMed: 15032552]
15. Ballester PJ, Mitchell JB. A Machine Learning Approach to Predicting Protein-Ligand Binding Affinity with Applications to Molecular Docking. *Bioinformatics.* 2010; 26:1169–75. [PubMed: 20236947]
16. Springer C, Adalsteinsson H, Young MM, Kegelmeyer PW, Roe DC. Postdock: A Structural, Empirical Approach to Scoring Protein Ligand Complexes. *J. Med. Chem.* 2005; 48:6821–31. [PubMed: 16250641]
17. Shoichet BK, Kuntz ID. Matching Chemistry and Shape in Molecular Docking. *Protein Eng.* 1993; 6:723–32. [PubMed: 7504257]
18. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput. Aided Mol. Des.* 1997; 11:425–45. [PubMed: 9385547]
19. Sato T, Honma T, Yokoyama S. Combining Machine Learning and Pharmacophore-Based Interaction Fingerprint for in Silico Screening. *J. Chem. Inf. Model.* 2010; 50:170–85. [PubMed: 20038188]
20. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* 2004; 47:1739–49. [PubMed: 15027865]
21. Durrant JD, McCammon JA. Nnscore: A Neural-Network-Based Scoring Function for the Characterization of Protein-Ligand Complexes. *J. Chem. Inf. Model.* 2010; 50:1865–71. [PubMed: 20845954]

22. Li Q, Wang Y, Bryant SH. A Novel Method for Mining Highly Imbalanced High-Throughput Screening Data in Pubchem. *Bioinformatics*. 2009; 25:3310–6. [PubMed: 19825798]
23. Tang Y, Zhang YQ, Chawla NV, Krasser S. Svms Modeling for Highly Imbalanced Classification. *IEEE Trans. Syst. Man. Cybern. B. Cybern.* 2009; 39:281–8. [PubMed: 19068445]
24. Tang, Y.; Zhang, Y-Q. Granular Svm with Repetitive Undersampling for Highly Imbalanced Protein Homology Prediction.. *Proceedings of 2006 IEEE International Conference on Granular Computing (IEEE-GrC2006)*; Atlanta. 2006. p. 457-460.
25. Guha R, Schurer SC. Utilizing High Throughput Screening Data for Predictive Toxicology Models: Protocols and Application to Mlscn Assays. *J. Comput. Aided. Mol. Des.* 2008; 22:367–84. [PubMed: 18283419]
26. Han L, Wang Y, Bryant SH. Developing and Validating Predictive Decision Tree Models from Mining Chemical Structural Fingerprints and High-Throughput Screening Data in Pubchem. *BMC Bioinformatics*. 2008; 9:401. [PubMed: 18817552]
27. Weis DC, Visco DP Jr, Faulon JL. Data Mining Pubchem Using a Support Vector Machine with the Signature Molecular Descriptor: Classification of Factor Xia Inhibitors. *J. Mol. Graph. Model.* 2008; 27:466–75. [PubMed: 18829357]
28. Zsoldos Z, Reid D, Simon A, Sadjad SB, Johnson AP. Ehits: A New Fast, Exhaustive Flexible Ligand Docking System. *J. Mol. Graph. Model.* 2007; 26:198–212. [PubMed: 16860582]
29. Lu XY, You QD, Chen YD. Recent Progress in the Identification and Development of Inha Direct Inhibitors of Mycobacterium Tuberculosis. *Mini. Rev. Med. Chem.* 2010; 10:181–92. [PubMed: 20408801]
30. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE. Drug Discovery Using Chemical Systems Biology: Repositioning the Safe Medicine Comtan to Treat Multi-Drug and Extensively Drug Resistant Tuberculosis. *PLoS Comput. Biol.* 2009; 5:e1000423. [PubMed: 19578428]
31. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. Bindingdb: A Web-Accessible Database of Experimentally Determined Protein-Ligand Binding Affinities. *Nucleic Acids Res.* 2007; 35:D198–201. [PubMed: 17145705]
32. Rucker C, Rucker G, Meringer M. Y-Randomization and Its Variants in Qspr/Qsar. *J. Chem. Inf. Model.* 2007; 47:2345–57. [PubMed: 17880194]
33. Huang N, Shoichet BK, Irwin JJ. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* 2006; 49:6789–801. [PubMed: 17154509]
34. Pham TA, Jain AN. Parameter Estimation for Scoring Protein-Ligand Interactions Using Negative Training Data. *J. Med. Chem.* 2006; 49:5856–68. [PubMed: 17004701]
35. He X, Alian A, Stroud R, Ortiz de Montellano PR. Pyrrolidine Carboxamides as a Novel Class of Inhibitors of Enoyl Acyl Carrier Protein Reductase from Mycobacterium Tuberculosis. *J. Med. Chem.* 2006; 49:6308–23. [PubMed: 17034137]
36. Rozwarski DA, Vilcheze C, Sugantino M, Bittman R, Sacchettini JC. Crystal Structure of the Mycobacterium Tuberculosis Enoyl-Acp Reductase, Inha, in Complex with Nad+ and a C16 Fatty Acyl Substrate. *J. Biol. Chem.* 1999; 274:15582–9. [PubMed: 10336454]
37. Kuo MR, Morbidoni HR, Alland D, Sneddon SF, Gourlie BB, Staveski MM, Leonard M, Gregory JS, Janjigian AD, Yee C, Musser JM, Kreiswirth B, Iwamoto H, Perozzo R, Jacobs WR Jr, Sacchettini JC, Fidock DA. Targeting Tuberculosis and Malaria through Inhibition of Enoyl Reductase: Compound Activity and Structural Data. *J. Biol. Chem.* 2003; 278:20851–9. [PubMed: 12606558]
38. Xie L, Bourne PE. A Robust and Efficient Algorithm for the Shape Description of Protein Structures and Its Application in Predicting Ligand Binding Sites. *BMC Bioinformatics*. 2007; 8(Suppl 4):S9. [PubMed: 17570152]
39. Xie L, Bourne PE. Detecting Evolutionary Relationships across Existing Fold Space, Using Sequence Order-Independent Profile-Profile Alignments. *Proc. Natl Acad. Sci. U S A.* 2008; 105:5441–6. [PubMed: 18385384]
40. Xie L, Bourne PE. A Unified Statistical Model to Support Local Sequence Order Independent Similarity Searching for Ligand-Binding Sites and Its Application to Genome-Based Drug Discovery. *Bioinformatics*. 2009; 25:i305–12. [PubMed: 19478004]

41. Kinnings SL, Xie L, Fung KH, Jackson RM, Bourne PE. The Mycobacterium Tuberculosis Drugome and Its Polypharmacological Implications. *PLoS Comput. Biol.* 2010; 6:e1000976. [PubMed: 21079673]
42. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28:235–42. [PubMed: 10592235]
43. 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 Deliv. Rev.* 2001; 46:3–26. [PubMed: 11259830]
44. Irwin JJ, Shoichet BK. Zinc--a Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* 2005; 45:177–82. [PubMed: 15667143]
45. Freundlich JS, Wang F, Vilcheze C, Gulten G, Langley R, Schiehser GA, Jacobus DP, Jacobs WR Jr, Sacchettini JC. Triclosan Derivatives: Towards Potent Inhibitors of Drug-Sensitive and Drug-Resistant Mycobacterium Tuberculosis. *ChemMedChem.* 2009; 4:241–8. [PubMed: 19130456]
46. Ye Y, Godzik A. Flexible Structure Alignment by Chaining Aligned Fragment Pairs Allowing Twists. *Bioinformatics.* 2003; 19(Suppl 2):ii246–55. [PubMed: 14534198]
47. Joachims T. Making Large-Scale Svm Learning Practical. :1999. MIT-Press.
48. Kirchmair J, Distinto S, Markt P, Schuster D, Spitzer GM, Liedl KR, Wolber G. How to Optimize Shape-Based Virtual Screening: Choosing the Right Query and Including Chemical Information. *J. Chem. Inf. Model.* 2009
49. Irwin JJ. Community Benchmarks for Virtual Screening. *J. Comput. Aided Mol. Des.* 2008; 22:193–9. [PubMed: 18273555]
50. Good AC, Oprea TI. Optimization of Camd Techniques 3. Virtual Screening Enrichment Studies: A Help or Hindrance in Tool Selection? *J. Comput. Aided Mol. Des.* 2008; 22:169–78. [PubMed: 18188508]
51. Clark RD, Webster-Clark DJ. Managing Bias in Roc Curves. *J. Comput. Aided Mol. Des.* 2008; 22:141–6. [PubMed: 18256892]
52. Kinnings SL, Jackson RM. Ligmach: A Multiple Structure-Based Ligand Matching Method for 3d Virtual Screening. *J. Chem. Inf. Model.* 2009; 49:2056–66. [PubMed: 19685924]
53. Cheeseright TJ, Mackey MD, Melville JL, Vinter JG. Fieldscreen: Virtual Screening Using Molecular Fields. Application to the Dud Data Set. *J. Chem. Inf. Model.* 2008; 48:2108–17. [PubMed: 18991371]

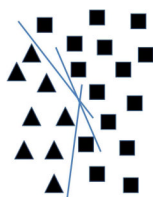


Figure 1.
The transformation of a linear classifier into a nonlinear classifier using a multiple-planar classifier.

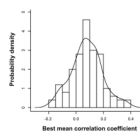


Figure 2. A histogram showing the distribution of the best mean correlation coefficients for 100 y-randomization trials when training the regression SVM

The highest mean correlation coefficient from five-fold cross-validation achieved across the 128 combinations of features is reported for each y-randomization.

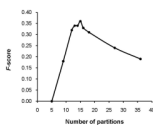


Figure 3. The F -scores associated with partitioning the decoys into different numbers of subsets when training the classification SVM

The highest F -score resulted from partitioning the decoys into 15 subsets.

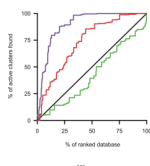


Figure 4. A ROC curve comparing the performance of eHiTS²⁸ (green line) with that of the SVM classification model (purple line) and that of the SVM regression model (red line) at discriminating actives from decoys in the InhA DUD dataset
Random performance is indicated by the black line.

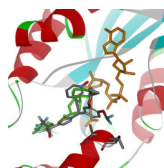
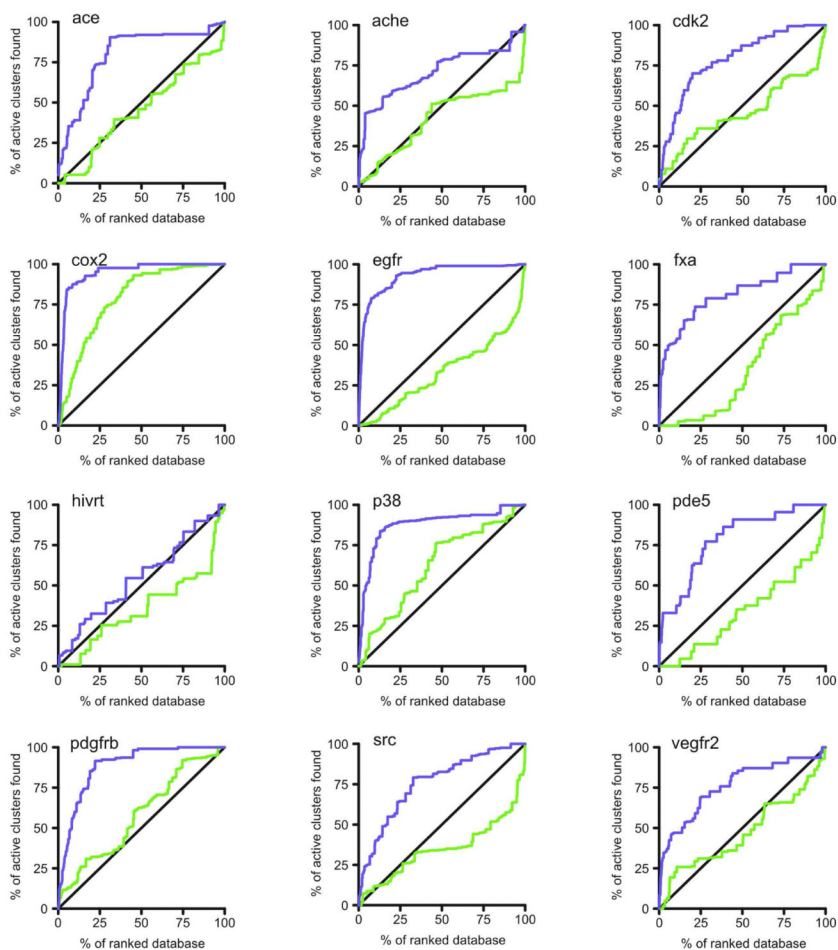


Figure 5. The binding pose of L-869298 (colored by element) in the InhA substrate binding site, as predicted by eHiTS²⁸

The NAD co-factor is shown in orange and a known high affinity direct InhA inhibitor, (3S)-1-cyclohexyl-N-(3,5-dichlorophenyl)-5-oxopyrrolidine-3-carboxamide (PDB code: 2h7m, PDB ligand code: 641, IC₅₀: 390nM), is shown in green. Potential hydrogen bonds between a fluorine atom of L-869298 and the Leu197 and Ala198 residues of InhA (shown in stick form, colored by element) are shown as green dashed lines.



ROC curves to compare the performance of the eHiTS scoring function and classification SVM across 12 selected DUD targets

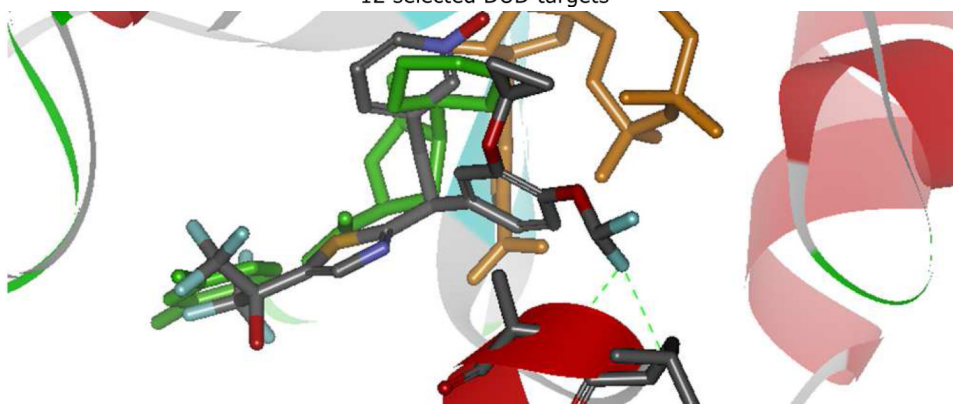


Figure 6. ROC curves to compare the performance of the eHiTS scoring function and classification SVM across 12 selected DUD targets

The lines are coloured as follows; eHiTS (green), classification SVM (purple), and random performance (black). Target abbreviations are as follows; ace, angiotensin-converting enzyme; ache, acetylcholinesterase; cdk2, cyclin-dependent kinase 2; cox2, cyclooxygenase-2; egfr, epidermal growth factor receptor; fxa, factor Xa; hivrt, HIV reverse

transcriptase; p38, P38 mitogen activated protein kinase; pde5, phosphodiesterase 5; pdgfrb, platelet derived growth factor receptor kinase; src, tyrosine kinase SRC; vegfr2, vascular endothelial growth factor receptor.

Table 1
The Spearman's rank correlation coefficients and *F*-scores resulting from the removal of different compulsory energy terms from the regression and classification models, respectively

In each case, the results are ordered by ascending performance, therefore the energy terms determined to make the greatest contribution towards the overall binding affinity are listed first. The thick horizontal lines indicate the turning point between energy terms whose removal had a significant effect on the performance, and those whose removal did not.

Regression model		Classification model	
Removed energy term	Spearman's rank correlation coefficient (ρ)	Removed energy term	<i>F</i> -score
pi-pi stack	0.058	van der Waals steric	0.000
solvation	0.062	solvation	0.019
hydrogen bonding	0.110	hydrophobic	0.020
hydrophobic	0.158	pi-pi stack	0.040
van der Waals steric	0.245	van der Waals strain	0.134
entropy	0.453	hydrogen bonding	0.166
electrostatics	0.475	electrostatics	0.171
van der Waals strain	0.486	entropy	0.187
other	0.627	other	0.206
none removed	0.637	none removed	0.232

Table 2

Three PDE5 inhibitors with the potential to inhibit InhA

SMAP *P*-values indicating the binding site similarity are given, as well as the docking scores of the PDE5 inhibitors with InhA, and their significance.

PDE5 inhibitor	Best SMAP <i>P</i> -value with InhA	eHiTS energy score rank among 303 PDE inhibitors	Best eHiTS energy score with InhA (pKi)	Mean of 500 energy scores (pKi)	Std. dev. of 500 energy scores (pKi)	Z-score
sildenafil (VIA)	3.99e-6	299	-9.55	-9.3080	1.2926	-0.1872
tadalafil (CIA)	4.36e-8	258	-11.74	-9.5229	1.7235	-1.2864
vardenafil (VDN)	2.35e-6	83	-11.74	-8.9159	1.3851	-2.0389

Table 3
The 10 PDE inhibitors with the highest predicted IC₅₀ values, as predicted by the SVM regression model

The names of the PDE inhibitors are given along with their actual targets and original energy scores with InhA, as predicted by eHiTS²⁸.

Ranking from regression model	Predicted IC ₅₀ from regression model (nM)	PDE inhibitor	PDE targets in BindingDB	Original eHiTS energy score (pKi)
1	8	phthalazinone, 22	PDE4a, PDE4b, PDE4d	-11.74
2	16	L-869299	PDE4a, PDE4b, PDE4c, PDE4d	-11.74
3	30	L-869298	PDE4a, PDE4b, PDE4c, PDE4d	-11.74
4	37	phthalazinone, 21	PDE4a, PDE4b, PDE4d	-9.91
5	61	V11294 analog 5u	PDE4a, PDE4b, PDE4d	-11.74
6	67	V11294 analog 5t	PDE4a, PDE4b, PDE4d	-11.74
7	77	V11294 analog 5j	PDE4a, PDE4b, PDE4d	-11.74
8	98	triazolothiadiazine, 8	PDE4a	-9.83
9	106	V11294 analog 5o	PDE4a, PDE4b, PDE4d	-12.01
10	138	V11294 analog 5s	PDE4a, PDE4b, PDE4d	-11.74

Table 4
The 10 most highly ranked PDE inhibitors, as predicted by the SVM classification model

The names of the PDE inhibitors are given along with their actual targets and original energy scores with InhA, as predicted by eHiTS²⁸.

Ranking from SVM classification model	PDE inhibitor	PDE targets in BindingDB	Original eHiTS energy score (pKi)
1	L-869299	PDE4a, PDE4b, PDE4c, PDE4d	-11.74
2	L-869298	PDE4a, PDE4b, PDE4c, PDE4d	-11.74
3	dimethoxyquinazoline derivative	PDE10a	-9.90
4	triazolothiadiazine, 8	PDE4a	-9.83
5	triazolothiadiazine, 11E	PDE4a	-11.74
6	triazolothiadiazine, 11G	PDE4a	-11.68
7	phthalazinone, 22	PDE4a, PDE4b, PDE4d	-11.74
8	V11294 analog 5u	PDE4a, PDE4b, PDE4d	-11.74
9	sildenafil 5-methyl alcohol analog 55	PDE5, PDE6	-9.50
10	V11294 analog 5t	PDE4a, PDE4b, PDE4d	-11.74