

Supplementary Data: Small-Molecule Inhibitor Starting Points Learned From Protein-Protein Interaction Inhibitor Structure

David Ryan Koes and Carlos J. Camacho

Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA 15260, USA

ABSTRACT

In this supplemental information we detail our SVM training methodology, describe the properties of the training set in more detail, and illustrate the potential application of SMISPs predictions to several targets of therapeutic importance. Additionally, we provide the cross-validation performance of the SVM classifier for various score thresholds (Table 1) and the performance of single attribute rules (Table 2) and two-attribute rules (Table 3). Three SMISPs predictions not shown in the primary paper due to homology are shown in Figure 1. Source code for the exhaustive rule learner, the benchmark SMISPs used for training, the training examples, a listing of the top rated predicted SMISPs for all non-redundant PPIs, and a ranking of PPIs by predicted small-molecule inhibitor susceptibility are included in a separate tar file.

SVM CLASSIFIER TRAINING

We train our SVM classifier using libSVM following the procedure suggested by Hsu *et al.* (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>): data values are normalized to have a zero mean and unit variance, a radial basis function is used as the SVM kernel, and a grid search is performed to identify the parameters (C and γ) that maximize the cross-validation accuracy. We found that a value of 2.0 for C and .000030518 for γ provided the best cross-validation accuracy and AUC (area under the ROC curve). The pairwise coupling method (Wu *et al.*, 2004) is used to generate probability estimates that are used to score potential SMISPs. The SVM training workflow was implemented as a process in RapidMiner 5.1 (<http://rapid-i.com/>).

The average and cumulative receiver operating characteristic (ROC) curves across all 39 cross-validation sets is shown in Figure 4. The average curve is indicative of the ability of the classifier to distinguish SMISPs from non-SMISPs within a single PPI, while the cumulative curve indicates how well the classifier would perform when evaluating residue clusters across different PPIs (which is not the expected usage of the classifier). The difference between the two curves indicates that different structures have different distributions of cluster scores. For example, in one structure the best scoring clusters may have a score of .8, while in another structure the best clusters have a score of .95 and the clusters with a score of .8 are true negatives.

TRAINING SET CHARACTERIZATION

The categorization of the 39 PPIs of the training set is shown in Figure 5(a) with the categorization of the entire non-redundant set of PPIs shown in Figure 5(b) for reference. The categories are extracted directly from the PDB, which contains a single categorical label for each structure. To give a better idea of the span of biological function represented by the training set than can be provided by a single label, we show a tag-cloud of the associated GO-terms in Figure 6.

DESCRIPTION OF INHIBITOR SMISPS

Figure 1(a) shows an inhibitor of the p53/MDM2 anti-cancer target. The automatically identified SMISP includes experimentally verified hot spots of this interaction (Lin *et al.*, 1994). In fact, the shown small molecule was designed precisely to mimic these residues (Popowicz *et al.*, 2011), and consequently the hydrophobic and hydrogen bond interactions of the small molecule and SMISP are nearly identical. An identical p53 SMISP with similar overlap with the bound ligand is predicted for the homologous p53/MDM4 interaction and is shown in Figure 2(a).

The BIR3 domain of the X-linked inhibitor of apoptosis protein (XIAP) binds to the N-terminal of Caspase-9 and the second mitochondria-derived activator of caspases (Smac). Peptidomimetics of the Smac AVPI N-terminal sequence inhibit these interactions (Mastrangelo *et al.*, 2008). The predicted SMISPs for both Caspase-9, Figure 1(b), and Smac, Figure 2(b), identifies these N-terminal residues and almost perfectly overlays the functional groups of the designed inhibitor.

Figure 1(c) shows an inhibitor of the gp41 HIV envelope protein. As with p53/MDM2, the predicted SMISP corresponds to a set of residues that have been verified experimentally to dominate the affinity of the interaction (Chan *et al.*, 1998), and the shown inhibitor was designed to target the exact hydrophobic pocket delineated by the SMISP (Stewart *et al.*, 2010).

Figure 1(d) shows an inhibitor of the anti-cancer Bcl-xL/Beclin 1 target. The groove identified by the SMISP was the target for the structural design of the shown inhibitor (Bruncko *et al.*, 2007), and the hydrophobic residues of the SMISP directly overlap hydrophobic moieties on the inhibitor. The related Bcl-2/BaxBH3 is shown in Figure 2(c) and is similar to the training SMISP of Bcl-xL/BaxBH3.

Figure 1(e) shows that in the predicted SMISP for HIV Integrase/p75 the carboxylate group of the aspartic acid almost

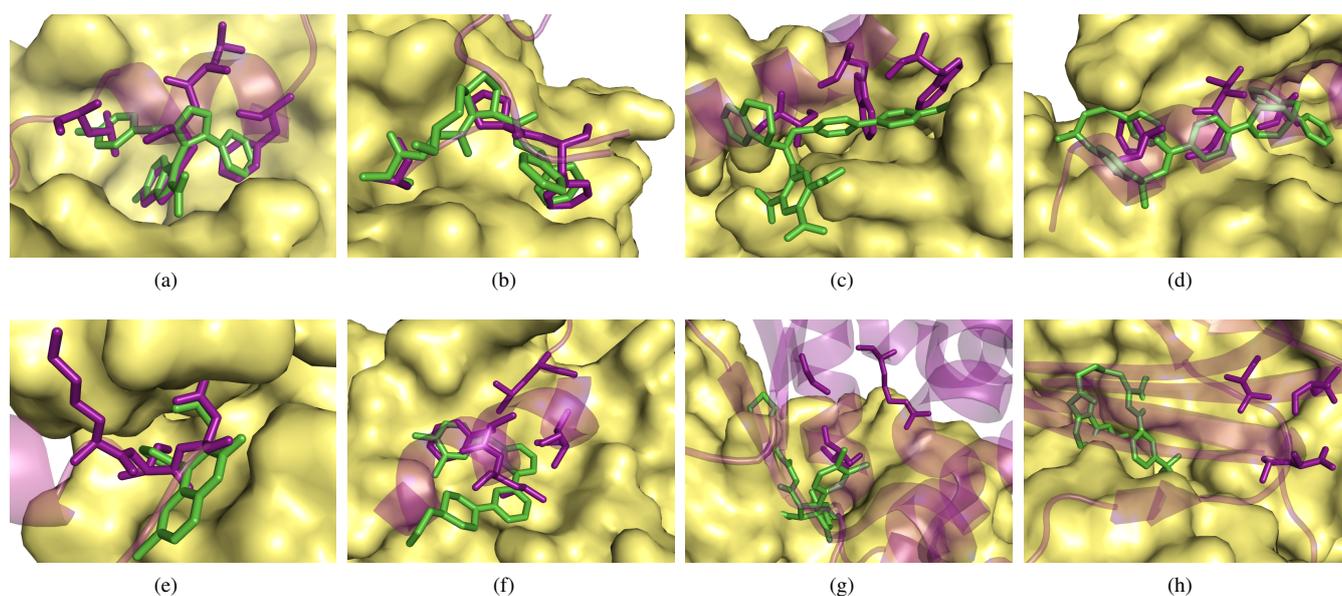


Fig. 1. Reproduction of Figure 2 from paper for convenience. SMISPs predictions for some of the PPIs from Table 1. The PPI is represented by a receptor protein (surface) and a ligand protein (transparent magenta). A small-molecule inhibitor (green) is posed by aligning the corresponding receptors. The single largest SMISP ranked in the top three is shown as magenta sticks. PDB access codes are provided in Table 1. In Figures (a-f) the predicted SMISPs overlap the inhibitor and at least partially delineate the binding pocket(s). In Figures (g-h) the SMISPs only marginally overlap the inhibitor and identify a nearby, but distinct, binding pocket. (a) p53/MDM2. (b) XIAP-BIR3/Caspase-9. (c) HIV gp41. (d) Bcl-xL/Beclin 1. (e) HIV-1 Integrase/p75. (f) ZipA/FtsZ. (g) HPV E1/E2. (h) TNF- α .

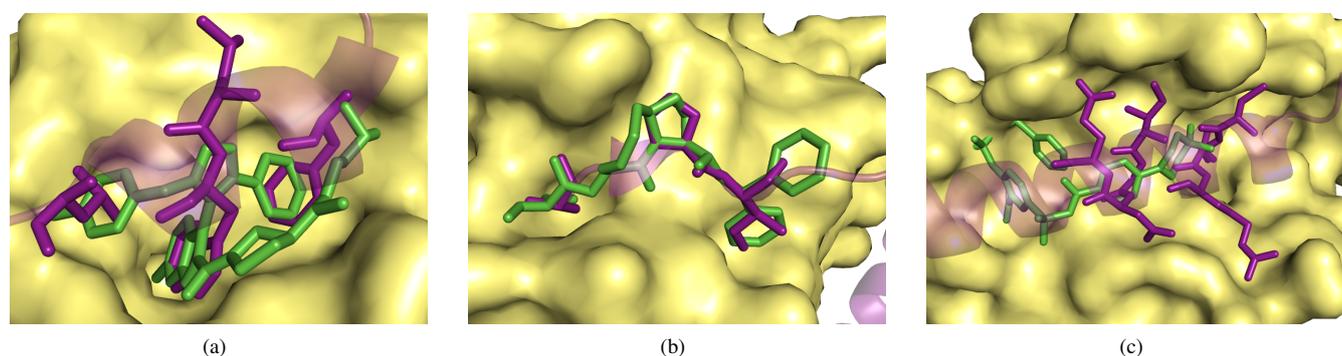


Fig. 2. SMISPs predictions for three validation set PPIs not shown in Figure 1. The PPI is represented by a receptor protein (surface) and a ligand protein (transparent magenta). A small-molecule inhibitor (green) is posed by aligning the corresponding receptors. The single largest SMISP ranked in the top three is shown as magenta sticks. (a) p53/MDM4. (b) XIAP-BIR3/Smac. (c) Bcl-2/BaxBH3.

perfectly overlays the identical group in the bound ligand, while the isoleucine fills the same hydrophobic pocket as a phenyl group on the ligand. In fact, these interactions were the starting point for the rational design of this inhibitor (Christ *et al.*, 2010).

Figure 1(f) shows an inhibitor of the anti-bacterial ZipA/FtsZ target. The four hydrophobic residues of the SMISP identify the hydrophobic region of the ZipA surface that is the target of an inhibitor found via a high-throughput screen (Rush III *et al.*, 2005) and the phenylalanine of the SMISP closely matches the phenyl ring of the inhibitor.

In the HPV E1/E2 complex of Figure 1(g), although there is some overlap between the predicted SMISP and the small molecule, most

of the inhibitor populates a groove that is not present in the PPI structure. In Figure 1(h), the TNF- α complex, the inhibitor is buried in a groove that is larger than the corresponding groove of the PPI.

POTENTIAL THERAPEUTIC SMISPs

To further illustrate the potential value of predicted SMISPs in structure-based design, we investigated several highly ranked PPI structures that are implicated in disease (the deposited structure includes the keywords disease, infection, cancer, or oncogene). Three examples selected for diversity of targets and therapeutic

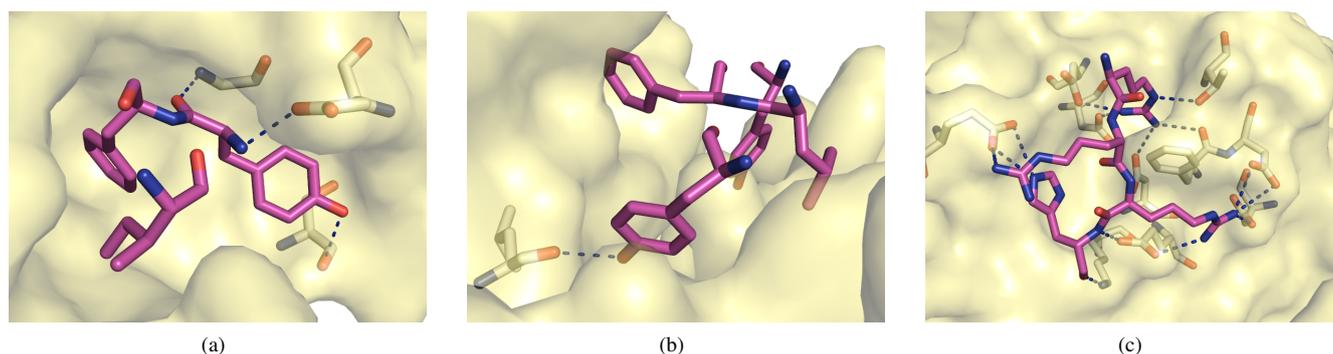


Fig. 3. SMISPs predictions for disease related PPI complexes. The receptor protein (surface) is shown with the largest SMISP of the three most highly ranked SMISPs for the complex (magenta sticks, oxygen shown in red, nitrogen in blue). The secondary structure of the ligand protein is omitted for clarity. Hydrogen bonds identified by PyMol (<http://www.pymol.org>) are shown with blue dashes. All three SMISPs contain many favorable interactions for structure-based design. (a) The autotransporter adhesin head domain BpaA from *Burkholderia pseudomallei*, the bacteria that causes melioidosis, is a trimer. The SMISP predicted from the trimeric complex (PDB: 3LAA) consists of three buried hydrophobic residues that make three hydrogen bonds. (b) The KSHV virus is the main cause of Kaposi's sarcoma. The viral protein ks-vFLIP is thought to play a role in subverting the host transcriptional pathway by binding to the IKKy component of the IKK enzyme complex. The SMISP predicted from the heteromeric complex (PDB: 3CL3) buries four residues of IKKy and forms a hydrogen bond with the backbone of ks-vFLIP. (c) PIM kinases have been implicated in many human cancers and are an active target for inhibitor design. The SMISP predicted from the complex of human PIM1 and a peptide substrate (PDB: 2C3I) contains three arginine residues and has a multitude of charge-charge and hydrogen bond interactions.

relevance are shown in Figure 3. The examples include a bacterial, viral, and anti-cancer target with homomeric, heteromeric, and protein-peptide structures.

Septicemic melioidosis, caused by the bacteria *Burkholderia pseudomallei*, has a greater than 10% mortality rate even when treated (Łyskowski *et al.*, 2011). The autotransporter adhesin head domain BpaA from *Burkholderia* is an obligate homotrimeric complex that plays an important role in host cell infection. Disrupting the formation of this complex with a small-molecule inhibitor might result in an effective alternative treatment. The SMISP, shown in Figure 3(a), predicted from the trimer (PDB: 3LAA), has a score of 0.99 and 17% of the interface clusters of the complex are identified as potential SMISPs.

The KSHV virus, the main cause of Kaposi's sarcoma, is one of several viruses that subvert the nuclear factor- κ B (NF- κ B) transcriptional pathway (Bagn eris *et al.*, 2008). The viral protein ks-vFLIP binds to the IKKy component of IKK enzyme complex which is part of the upstream NF- κ B signal transduction cascade. An inhibitor of this interaction might prevent tumorigenesis. The top ranked SMISP, shown in Figure 3(b), predicted from the heteromeric structure (PDB: 3CL3), has a score of 0.95 and 58% of the interface clusters of the complex are identified as potential SMISPs.

PIM kinases have been implicated in many human cancers and are an active target for inhibitor design (Nawijn *et al.*, 2011). The SMISP, shown in Figure 3(c), predicted from the complex of human PIM1 and a peptide substrate (PDB: 2C3I), has a score of 0.89 and 49% of the interface clusters of the complex are identified as potential SMISPs.

In all three cases, the predicted SMISPs identify many potential interactions, such as hydrogen bonds, charge-charge interactions, and hydrophobic contacts, that can be explored through structure-based design. For example, software such as LigandScout (Wolber and Langer, 2004) or Pharmer (Koes and Camacho, 2011)

can automatically identify the interaction pharmacophore, the spatial arrangement of essential features of the interaction. This pharmacophore can then be used to quickly screen large compound libraries to produce an enriched subset of compounds for further investigation.

Table 1. Leave-one-complex-out cross-validation performance of the SVM classifier. Performance values, defined in Table 4, are averages across the 39 train-test cross-validation sets and are shown with the standard error of the mean for different score thresholds.

Threshold	Accuracy	Precision	Recall	Specificity	F1
0.50	72±3.7%	62±6.1%	71±7.0%	73±5.3%	64±6.3%
0.55	74±3.7%	64±6.3%	71±7.1%	77±5.3%	64±6.5%
0.60	72±3.8%	61±6.7%	62±7.7%	81±5.1%	57±7.1%
0.65	68±3.6%	67±6.8%	50±7.8%	87±4.4%	47±7.2%
0.70	64±3.4%	65±7.0%	40±7.5%	89±4.3%	38±7.0%
0.75	62±3.2%	66±7.0%	33±7.3%	91±3.5%	32±6.9%

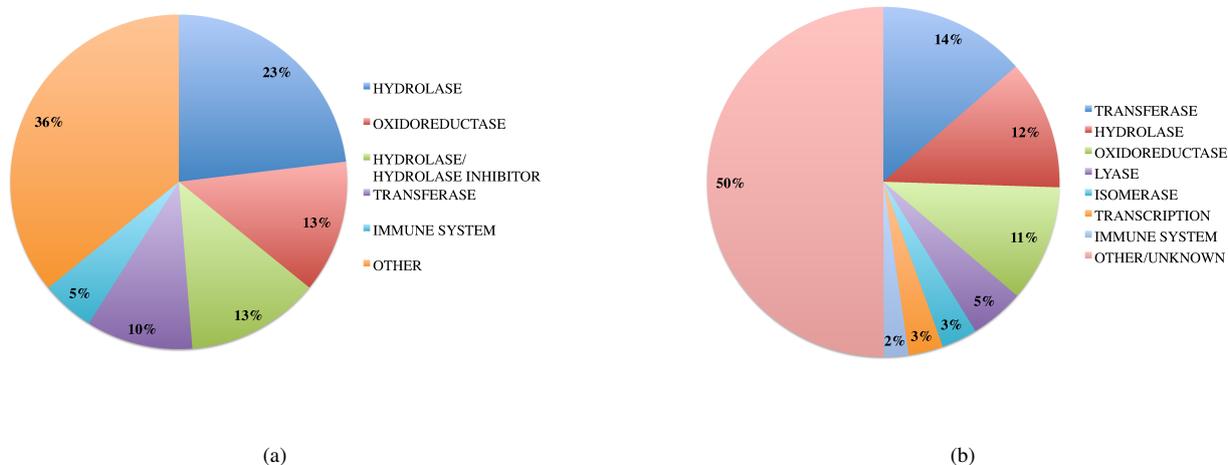


Fig. 5. The distribution of PDB category assignments within the training set (a) and the entire non-redundant set (b).

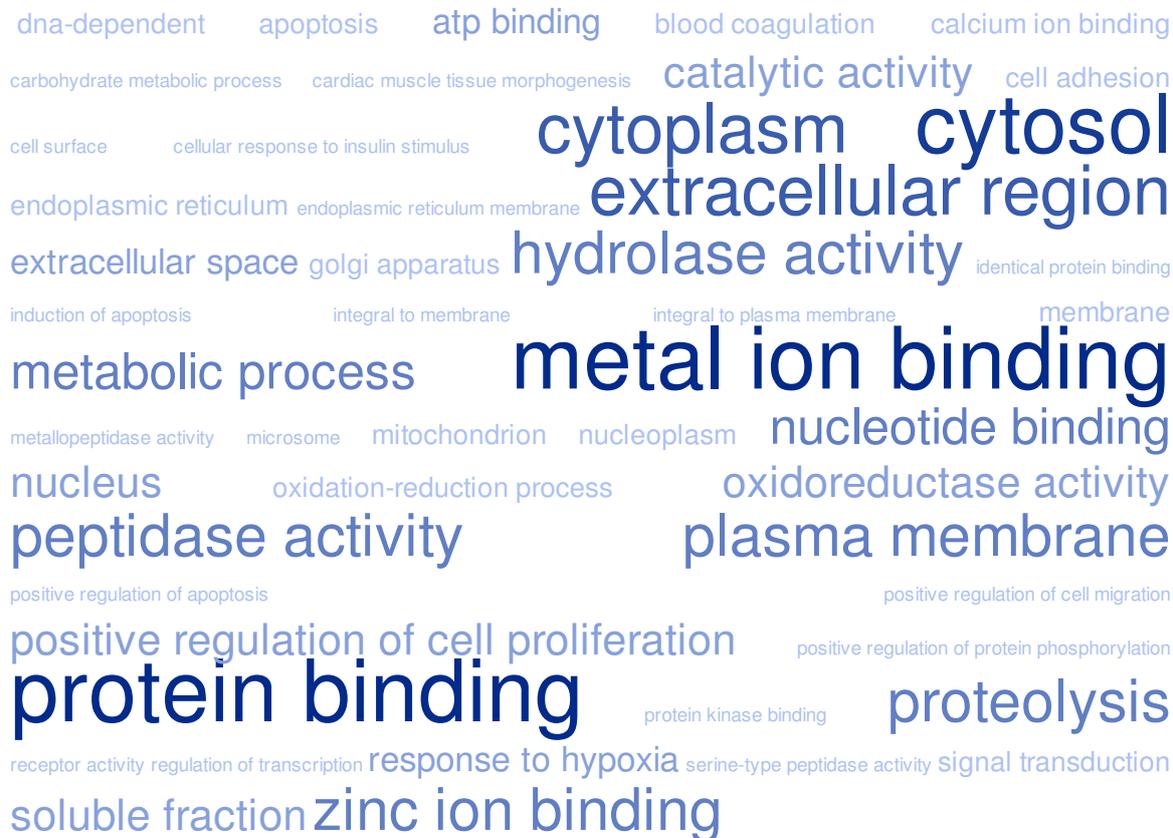


Fig. 6. A tag-cloud of the gene ontology (GO) terms associated with all the complexes in the training set.

Table 2. Single attribute rule classifiers. The optimal single attribute rules for each of the computed properties are shown ranked by information gain. The information gain and rule thresholds are computed using the entire training set. The performance values, defined in Table 4, are averages across the 39 train-test cross-validation sets and are shown with the standard error of the mean.

Info. Gain	Rule	Accuracy	Precision	Recall	Specificity	F1
0.137	Ave Δ SASA \geq 44.6	68 \pm 3.6%	57 \pm 5.5%	73 \pm 6.8%	63 \pm 4.8%	63 \pm 5.8%
0.128	Ave Δ SASA% \geq 39.6	68 \pm 3.7%	57 \pm 5.9%	69 \pm 7.0%	67 \pm 5.1%	61 \pm 6.1%
0.124	Max Δ SASA \geq 89.5	69 \pm 4.1%	53 \pm 6.6%	64 \pm 7.5%	74 \pm 5.3%	57 \pm 6.9%
0.113	Min Δ SASA% \geq 3.9	58 \pm 2.6%	50 \pm 4.0%	78 \pm 6.2%	38 \pm 4.5%	60 \pm 4.5%
0.113	Total Δ SASA \geq 181	66 \pm 4.0%	43 \pm 7.0%	51 \pm 7.9%	81 \pm 5.2%	46 \pm 7.2%
0.103	Ave Δ G ^{FC} $<$ -2.3	67 \pm 3.9%	47 \pm 7.1%	51 \pm 7.7%	83 \pm 3.8%	48 \pm 7.2%
0.100	Min Δ SASA \geq 3.66	61 \pm 2.2%	57 \pm 2.1%	94 \pm 3.3%	29 \pm 4.1%	70 \pm 2.3%
0.099	Total Δ SASA% \geq 122	67 \pm 3.7%	50 \pm 6.3%	65 \pm 7.6%	69 \pm 5.5%	56 \pm 6.7%
0.081	Ave Δ G ^{FC} _{dsolv} $<$ -1.64	61 \pm 3.4%	29 \pm 6.8%	30 \pm 7.2%	92 \pm 1.7%	29 \pm 6.8%
0.077	Min Δ G ^{FC} _{dsolv} $<$ -2.9	61 \pm 3.8%	35 \pm 6.9%	38 \pm 7.7%	84 \pm 3.3%	36 \pm 7.1%
0.070	Max Δ Δ G ^R \geq 0.427	62 \pm 3.1%	50 \pm 4.6%	78 \pm 6.5%	47 \pm 5.4%	61 \pm 5.2%
0.070	Max Δ SASA% \geq 51.6	51 \pm 2.8%	28 \pm 5.1%	44 \pm 7.9%	58 \pm 5.7%	34 \pm 6.1%
0.066	Total Δ G ^{FC} _{dsolv} $<$ -5.69	58 \pm 3.0%	22 \pm 6.2%	22 \pm 6.5%	93 \pm 1.9%	22 \pm 6.2%
0.065	Total Δ G ^{FC} $<$ -6.28	55 \pm 2.8%	26 \pm 5.9%	30 \pm 7.0%	80 \pm 4.1%	27 \pm 6.0%
0.059	Min Δ G ^{FC} $<$ -3.43	56 \pm 3.2%	30 \pm 5.8%	41 \pm 7.8%	70 \pm 4.8%	34 \pm 6.5%
0.052	Max Δ G ^{FC} _{dsolv} $<$ -1.68	56 \pm 2.6%	15 \pm 5.6%	14 \pm 5.4%	98 \pm 0.7%	14 \pm 5.4%
0.046	Max Δ G ^{FC} $<$ -2.08	57 \pm 2.7%	21 \pm 6.4%	18 \pm 5.9%	95 \pm 1.9%	18 \pm 5.7%
0.042	Min Δ G ^{FC} _{elec} $<$ -1.3	61 \pm 3.7%	44 \pm 5.6%	64 \pm 7.7%	58 \pm 4.8%	52 \pm 6.4%
0.041	Total Δ G ^{FC} _{elec} $<$ -15.4	54 \pm 2.3%	13 \pm 5.1%	13 \pm 5.2%	95 \pm 3.1%	13 \pm 4.9%
0.039	Max Δ G ^{FC} _{elec} $<$ -0.17	60 \pm 3.3%	36 \pm 6.9%	36 \pm 7.3%	84 \pm 3.4%	34 \pm 6.6%
0.038	Ave Δ Δ G ^R \geq 0.257	59 \pm 4.2%	44 \pm 5.9%	57 \pm 7.6%	62 \pm 4.7%	49 \pm 6.5%
0.038	Total Rate4 \geq 1.6	52 \pm 2.7%	44 \pm 3.9%	80 \pm 6.5%	24 \pm 5.3%	57 \pm 4.7%
0.037	Max Rate4 \geq 0.995	49 \pm 2.5%	45 \pm 3.1%	85 \pm 5.6%	14 \pm 3.3%	59 \pm 3.9%
0.034	Total Δ Δ G ^R \geq 0.405	54 \pm 4.0%	38 \pm 5.5%	53 \pm 7.7%	56 \pm 4.8%	44 \pm 6.3%
0.033	Min Rate4 \geq 0.301	55 \pm 2.5%	48 \pm 3.7%	85 \pm 5.8%	25 \pm 5.5%	60 \pm 4.3%
0.032	Min Cons $<$ 0.086	52 \pm 1.5%	4 \pm 3.1%	5 \pm 3.6%	99 \pm 0.9%	5 \pm 3.3%
0.031	Ave Δ G ^{FC} _{elec} $<$ -1.46	51 \pm 2.6%	16 \pm 4.8%	23 \pm 6.8%	79 \pm 4.3%	19 \pm 5.6%
0.030	Ave Rate4 \geq 0.262	44 \pm 2.2%	35 \pm 3.8%	68 \pm 7.5%	21 \pm 5.4%	46 \pm 5.0%
0.023	Total Cons $<$ 0.804	51 \pm 0.7%	9 \pm 3.3%	14 \pm 5.5%	87 \pm 4.4%	11 \pm 4.0%
0.021	Min Δ Δ G ^R \geq 0.44	47 \pm 1.2%	3 \pm 2.3%	2 \pm 1.4%	92 \pm 2.2%	2 \pm 1.7%
0.021	Ave Cons $<$ 0.325	49 \pm 1.3%	4 \pm 2.5%	8 \pm 4.3%	91 \pm 3.9%	6 \pm 3.2%
0.017	Max Cons $<$ 0.855	49 \pm 1.7%	19 \pm 4.1%	35 \pm 7.7%	63 \pm 7.1%	25 \pm 5.3%

Table 4. The derivation of various metrics of classification performance. Values are computed with respect to the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

Metric	Derivation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision (positive predictive value)	$\frac{TP}{TP+FP}$
Recall (sensitivity, hit rate)	$\frac{TP}{TP+FN}$
Specificity (true negative rate)	$\frac{TN}{TN+FP}$
F1 (harmonic mean of precision and recall)	$\frac{2TP}{2TP+FN+FP}$

REFERENCES

- Bagn ris, C., Ageichik, A. V., Cronin, N., Wallace, B., Collins, M., Boshoff, C., Waksman, G., and Barrett, T. (2008). Crystal structure of a vflipp-ikk γ complex: Insights into viral activation of the ikk signalosome. *Mol Cell*, **30**(5), 620–631.
- Bruncko, M., Oost, T. K., Belli, B. A., Ding, H., Joseph, M. K., Kunzer, A., Martineau, D., McClellan, W. J., Mitten, M., and Ng, S. C. (2007). Studies leading to potent, dual inhibitors of bcl-2 and bcl-xl. *Journal of medicinal chemistry*, **50**(4), 641–662.
- Chan, D., Chutkowski, C., and Kim, P. (1998). Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *PNAS*, **95**(26), 15613.
- Christ, F., Voet, A., Marchand, A., Nicolet, S., Desimmie, B. A., Marchand, D., Bardiot, D., Van der Veken, N. J., Van Remoortel, B., and Strelkov, S. V. (2010). Rational design of small-molecule inhibitors of the ledgf/p75-integrase interaction and hiv replication. *Nature Chemical Biology*, **6**(6), 442–448.
- Koes, D. R. and Camacho, C. J. (2011). Pharmer: Efficient and exact pharmacophore search. *Journal of Chemical Information and Modeling*, **51**(6), 1307–1314.

Table 3. The ten most informative two-attribute rule classifiers. The information gain and rule thresholds are computed using the entire training set. The performance values, defined in Table 4, are averages across the 39 train-test cross-validation sets and are shown with the standard error of the mean.

Info. Gain		Rule		Accuracy	Precision	Recall	Specificity	F1
0.175	Ave	Δ SASA	\geq 44.6	72 \pm 3.8%	61 \pm 6.3%	69 \pm 7.0%	74 \pm 5.1%	63 \pm 6.3%
	Ave	Δ SASA%	\geq 39.6					
0.171	Ave	Δ G ^{FC}	$<$ -2.27	71 \pm 3.8%	50 \pm 7.4%	51 \pm 7.7%	91 \pm 3.2%	49 \pm 7.4%
	Total	Δ SASA%	\geq 125					
0.167	Max	Δ Δ G ^R	\geq 0.425	66 \pm 3.8%	47 \pm 6.7%	54 \pm 7.7%	77 \pm 4.5%	49 \pm 6.9%
	Ave	Δ SASA	\geq 46.1					
0.165	Min	Δ G ^{FC}	$<$ -0.88	69 \pm 3.8%	58 \pm 6.0%	70 \pm 7.0%	68 \pm 4.8%	62 \pm 6.2%
	Ave	Δ SASA	\geq 44.6					
0.161	Ave	Δ SASA	\geq 44.6	70 \pm 3.7%	59 \pm 5.8%	73 \pm 6.8%	67 \pm 5.0%	64 \pm 6.0%
	Total	Δ SASA%	\geq 79.2					
0.160	Max	Δ G ^{FC}	$<$ 3.98	68 \pm 3.6%	56 \pm 6.0%	66 \pm 7.2%	70 \pm 4.1%	59 \pm 6.2%
	Ave	Δ SASA	\geq 44.6					
0.160	Max	Δ SASA	\geq 59.4	66 \pm 3.8%	49 \pm 6.4%	60 \pm 7.6%	72 \pm 5.1%	53 \pm 6.7%
	Ave	Δ SASA%	\geq 39.6					
0.159	Min	Δ G ^{FC}	$<$ -0.88	69 \pm 3.8%	57 \pm 6.3%	67 \pm 7.2%	72 \pm 4.7%	60 \pm 6.4%
	Ave	Δ SASA%	\geq 39.6					
0.158	Ave	Δ SASA	\geq 44.6	67 \pm 3.7%	54 \pm 5.9%	68 \pm 7.2%	66 \pm 4.6%	59 \pm 6.2%
	Min	Δ G ^{FC} _{elec}	$<$ 0.36					
0.158	Ave	Δ G ^{FC}	$<$ 1.14	67 \pm 3.7%	54 \pm 6.0%	68 \pm 7.2%	67 \pm 4.7%	59 \pm 6.3%
	Ave	Δ SASA	\geq 44.6					

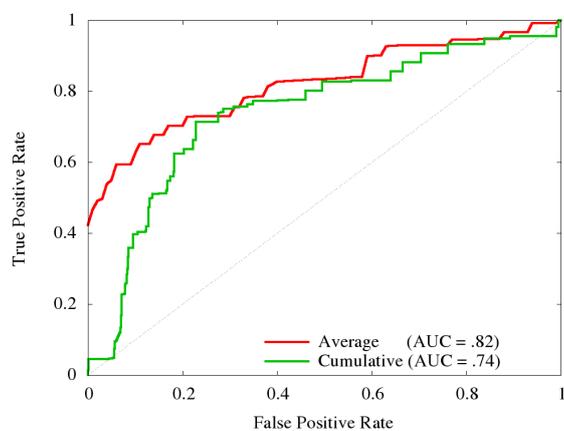


Fig. 4. The average and cumulative ROC curves of the cross-validation performance of the SVM classifier. The cross-validation generates 39 distinct prediction sets with distinct ROC curves. The average ROC simply takes the average of these 39 curves. The cumulative curve combines all 39 prediction sets into one set. The average curve indicates how well the classifier performs when classifying residues clusters from the same PPI (the expected usage of the classifier) while the cumulative curve is indicative of the performance when classifying clusters from different PPIs.

- Lin, J., Chen, J., Elenbaas, B., and Levine, A. J. (1994). Several hydrophobic amino acids in the p53 amino-terminal domain are required for transcriptional activation, binding to mdm-2 and the adenovirus 5 e1b 55-kd protein. *Genes Dev.* **8**(10), 1235.
- Lyskowski, A., Leo, J. C., and Goldman, A. (2011). Structure and biology of trimeric autotransporter adhesins. *Bacterial Adhesion*, pages 143–158.
- Mastrangelo, E., Cossu, F., Milani, M., Sorrentino, G., Lecis, D., Delia, D., Manzoni, L., Drago, C., Seneci, P., Scolastico, C., Rizzo, V., and Bolognesi, M. (2008). Targeting the x-linked inhibitor of apoptosis protein through 4-substituted azabicyclo[5.3.0]alkane smac mimetics. structure, activity, and recognition principles. *J. Mol. Biol.*, **384**(3), 673–689.
- Nawijn, M. C., Alendar, A., and Berns, A. (2011). For better or for worse: the role of pim oncogenes in tumorigenesis. *Nat Rev Cancer*, **11**(1), 23–34. 10.1038/nrc2986.
- Popowicz, G., Dömling, A., and Holak, T. (2011). The Structure-Based Design of Mdm2/Mdmx-p53 Inhibitors Gets Serious. *Angew. Chem. Int. Ed.*
- Rush III, T. S., Grant, J. A., Mosyak, L., and Nicholls, A. (2005). A shape-based 3-d scaffold hopping method and its application to a bacterial protein-protein interaction. *Journal of medicinal chemistry*, **48**(5), 1489–1495.
- Stewart, K. D., Huth, J. R., Ng, T. I., McDaniel, K., Hutchinson, R. N., Stoll, V. S., Mendoza, R. R., Matayoshi, E. D., Carrick, R., and Mo, H. M. (2010). Non-peptide entry inhibitors of hiv-1 that target the gp41 coiled coil pocket. *Bioorg. Med. Chem. Lett.*, **20**(2), 612–617.
- Wolber, G. and Langer, T. (2004). Ligandscout: 3-d pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Journal of Chemical Information and Modeling*, **45**(1), 160–169.
- Wu, T. F., Lin, C. J., and Weng, R. C. (2004). Probability estimates for multi-class classification by pairwise coupling. *JMLR*, **5**, 975–1005.