

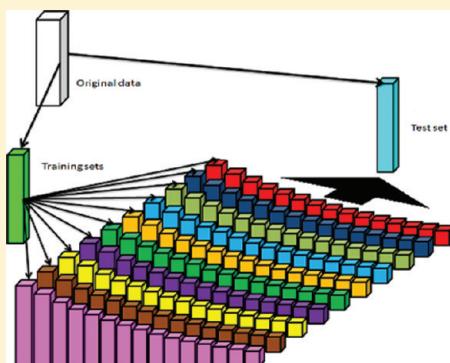
Rank Order Entropy: Why One Metric Is Not Enough

Margaret R. McLellan,[†] M. Dominic Ryan,[‡] and Curt M. Breneman^{*†}

[†]Department of Chemistry, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, New York 12180, United States

[‡]Cubist Pharmaceuticals, Inc., Lexington, Massachusetts, United States

ABSTRACT: The use of Quantitative Structure–Activity Relationship models to address problems in drug discovery has a mixed history, generally resulting from the misapplication of QSAR models that were either poorly constructed or used outside of their domains of applicability. This situation has motivated the development of a variety of model performance metrics (r^2 , PRESS r^2 , F-tests, etc.) designed to increase user confidence in the validity of QSAR predictions. In a typical workflow scenario, QSAR models are created and validated on training sets of molecules using metrics such as Leave-One-Out or many-fold cross-validation methods that attempt to assess their internal consistency. However, few current validation methods are designed to directly address the stability of QSAR predictions in response to changes in the information content of the training set. Since the main purpose of QSAR is to quickly and accurately estimate a property of interest for an untested set of molecules, it makes sense to have a means at hand to correctly set user expectations of model performance. In fact, the numerical value of a molecular prediction is often less important to the end user than knowing the rank order of that set of molecules according to their predicted end point values. Consequently, a means for characterizing the stability of predicted rank order is an important component of predictive QSAR. Unfortunately, none of the many validation metrics currently available directly measure the stability of rank order prediction, making the development of an additional metric that can quantify model stability a high priority. To address this need, this work examines the stabilities of QSAR rank order models created from representative data sets, descriptor sets, and modeling methods that were then assessed using Kendall Tau as a rank order metric, upon which the Shannon entropy was evaluated as a means of quantifying rank-order stability. Random removal of data from the training set, also known as Data Truncation Analysis (DTA), was used as a means for systematically reducing the information content of each training set while examining both rank order performance and rank order stability in the face of training set data loss. The premise for DTA ROE model evaluation is that the response of a model to incremental loss of training information will be indicative of the quality and sufficiency of its training set, learning method, and descriptor types to cover a particular domain of applicability. This process is termed a “rank order entropy” evaluation or ROE. By analogy with information theory, an unstable rank order model displays a high level of implicit entropy, while a QSAR rank order model which remains nearly unchanged during training set reductions would show low entropy. In this work, the ROE metric was applied to 71 data sets of different sizes and was found to reveal more information about the behavior of the models than traditional metrics alone. Stable, or consistently performing models, did not necessarily predict rank order well. Models that performed well in rank order did not necessarily perform well in traditional metrics. In the end, it was shown that ROE metrics suggested that some QSAR models that are typically used should be discarded. ROE evaluation helps to discern which combinations of data set, descriptor set, and modeling methods lead to usable models in prioritization schemes and provides confidence in the use of a particular model within a specific domain of applicability.



■ INTRODUCTION

Model validation is not a solved problem in Quantitative Structure–Activity Relationships modeling, though there are many techniques to validate models that enjoy varying degrees of success. These techniques are often particular to the task and often incorrectly applied to QSAR models. In prioritization schemes, often used in drug discovery, ensuring the stability of rank order predictions can be more important than the prediction of floating point values, especially in the selection of promising scaffolds. The stability of rank order prediction as a validation of model performance could therefore have significant utility on lead candidate prioritization schemes.

Model stability is of critical importance in determining their utility, as top-ranking predictions affect future decisions. Because model performance is dependent upon the size and quality of data sets as well as the parameters used in model creation, understanding how models perform in response to changes in the training set helps to establish the validity of the model within its domain of applicability. If models are highly sensitive to changes in the parameters used in the creation process, the trustworthiness of the combination of data, descriptors, and

Received: April 14, 2011

Published: August 29, 2011

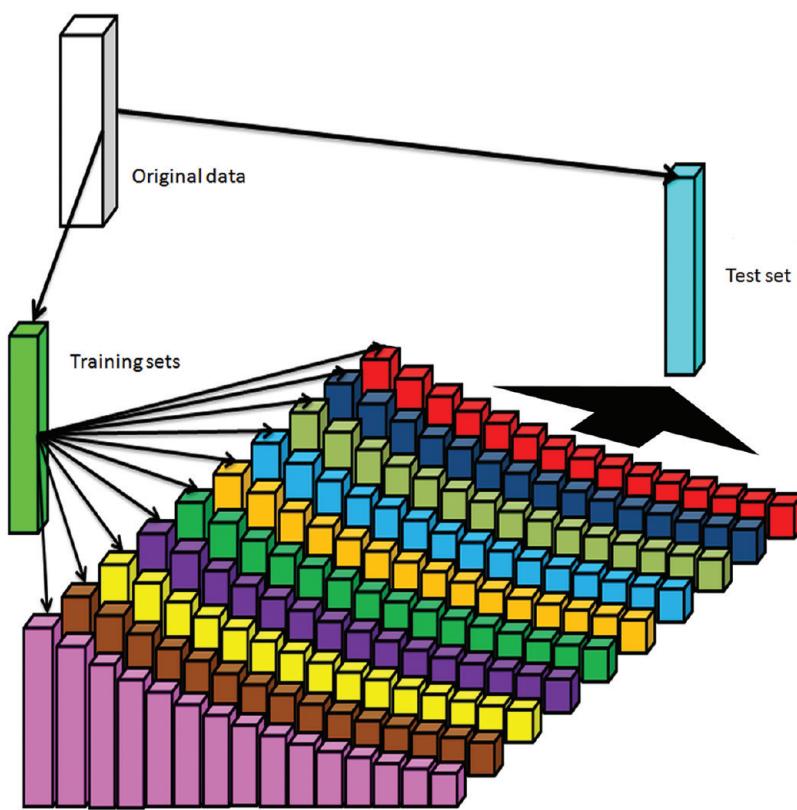


Figure 1. Data truncation in preparation for modeling: each time a data set is split, 151 training sets are created. The test set remains constant for these 151 truncations.

modeling method used to create the model must be questioned. Creating a stress test to reveal trustworthy combinations of data, descriptors, and modeling methods necessitates the use of validation metrics that also determine model rank order stability.

Evaluating rank order stability involves utilizing metrics that evaluate the usefulness of a model in making rank-order predictions. While many traditional model performance metrics exist including r^2 , Q^2 (the predicted residual sum of squares), and Root Mean Squared Error (RMSE), they assess the quality of floating point predictions and not rank order, which is of greater concern in discerning rank order stability.¹ Commonly used rank order metrics include Spearman's rank correlation coefficient and Kendall Tau.^{2,3} The main difference between them is that Spearman's rank correlation coefficient penalizes all changes in predicted rank order equally, while Kendall Tau is less stringent, and credits retention of portions of relative rank order despite any shift in predicted rank order. As retention of relative rank order has value within prioritization schemes, Kendall Tau more completely fulfills the requirements of this project. Evaluation of rank order performance alone by Kendall Tau does not necessarily validate model performance, unless the stability of the ranking results are also considered, resulting in the need for a metric (ROE) that quantifies the stability of a predicted rank order. Finally, it should be said that a favorable ROE result is not necessarily a definitive metric of model performance but can increase user confidence in predicted rankings.

Currently existing validation techniques mostly fall within two basic types: methods that affect model construction and methods that test model construction methods. Traditional validation techniques that influence model creation include Leave-One-Out and

bootstrapping methods. These model validation techniques incorporate multiple models into one, after which r^2 or other similar metrics are applied to the results. Such techniques are often used in combination with other validation methods, such as those that discern the potential for overfitting, but even this combination of validation methods often fails to provide an adequate indication of model stability. Validation metrics that directly address overfitting include y -scrambling and partial y -scrambling.^{4,5} Though y -scrambling assesses the propensity of a particular modeling method toward overfitting, it is only one indicator of potential prediction issues when the resulting models are used on unknown data sets.

Model validation can be extended by adding a rank order stability assessment step in prioritization schemes. Model stability can be thought of as disorder in model output in response to changes in training set information content and therefore may be assessed using metrics such as Shannon information entropy.⁶ Assessment of ranking stability requires determining changes in predicted rank order with respect to changes in the training data, during which multiple predictions from the decreasing data set are evaluated. As stated earlier, changes in model performance resulting from the reduction of information in the training data can be represented as disorder within the modeling results - a situation that can be enumerated using the concepts of information entropy. Shannon entropy, as described in literature, is characterized by its measurement of information entropy through examination of the distribution of a set of values over a domain.⁶ Evaluating model stability using the Shannon entropy of Kendall Tau requires the characterization of the range of Kendall Tau over the set of data set reductions.³ In order to

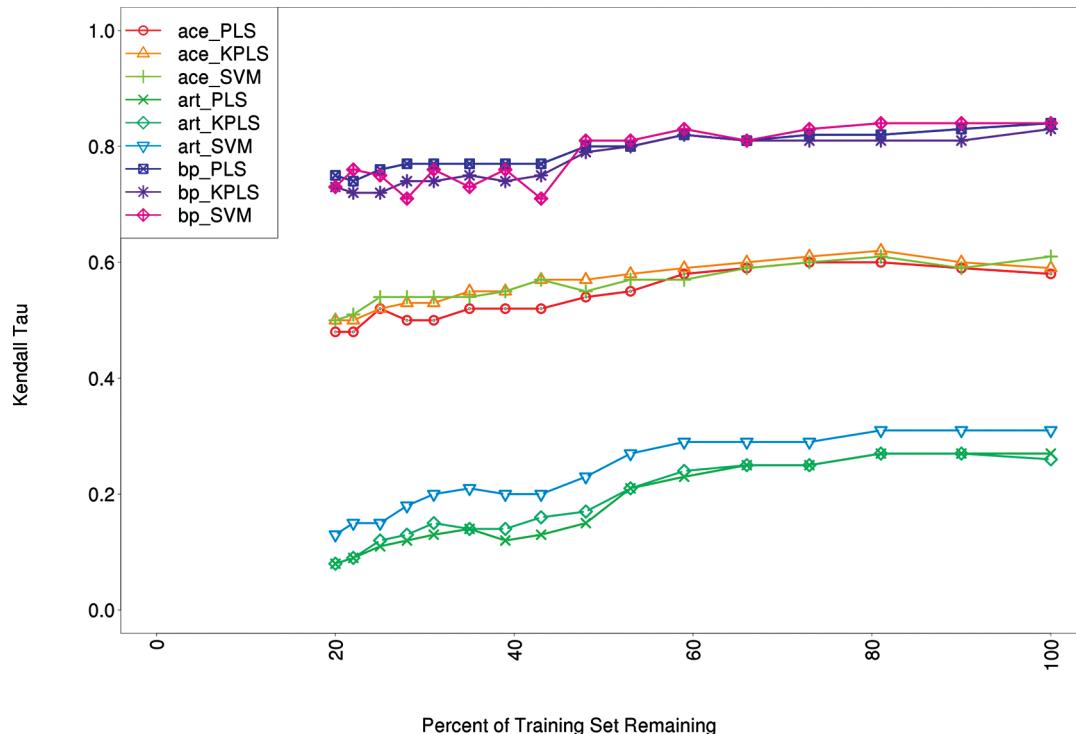


Figure 2. Data set evaluations without reduced truncations: ACE, artemisinin, and boiling point model behavior.

accomplish this, 40 bins of 0.05 width were created to allow for some minor variation within Kendall Tau during the Shannon entropy evaluation. Rank Order Entropy is determined by the population distribution of Kendall Tau across these bins. Rank Order Entropy analysis examines the variance in rank within a model as a data truncation analysis (DTA) is performed. This process corresponds to removing information from the training data by removing cases in a regular fashion, akin to taking a derivative of the behavior of the modeling results with respect to information loss. To reduce bias, training data removals are random and performed multiple times. The data truncation analysis performed within ROE analysis is represented in Figure 1.

The goal of Rank Order Entropy analysis is to evaluate whether a particular combination of data, descriptors, and modeling method can make robust predictions in the face of fewer data points, representing a decrease in training information. Stable combinations of descriptors, modeling method, and data result in models that retain the same level of rank order prediction even without the majority of the training data. The application of rank order metrics therefore provides a means for determining the reliability of a particular combination of data, descriptors, and modeling method. A stable rank order is the key outcome in determining whether a model can be trusted for a given task.

MATERIALS AND METHODS

There are multiple steps involved in performing a ROE assessment: data truncation, model creation, rank order evaluation, and stability evaluation. ROE evaluation is available as an online toolkit at <http://reccr.chem.rpi.edu/Software/ROE/ROE-index.html>.

Data truncation begins by creating initial training and testing sets from the original data set. This training/testing split is produced by first sorting the cases within the data set by activity and then splitting every other entry into either training or testing. This sorting and splitting is done in an effort to create a test set that is highly representative of the training set. There is an option to repeat the splitting process by performing it randomly, which creates another training/testing set pair considered separately from the original even/odd split. The test set remains constant after the initial splitting for modeling with the corresponding training sets to provide a stable test set for model performance evaluation. Once the test set is created, the training set truncation process can begin. At this point, the training set is randomly reduced in size by ten percent per iteration for ten iterations. This cycle is repeated fourteen times, creating 151 training sets, as shown in Figure 1.

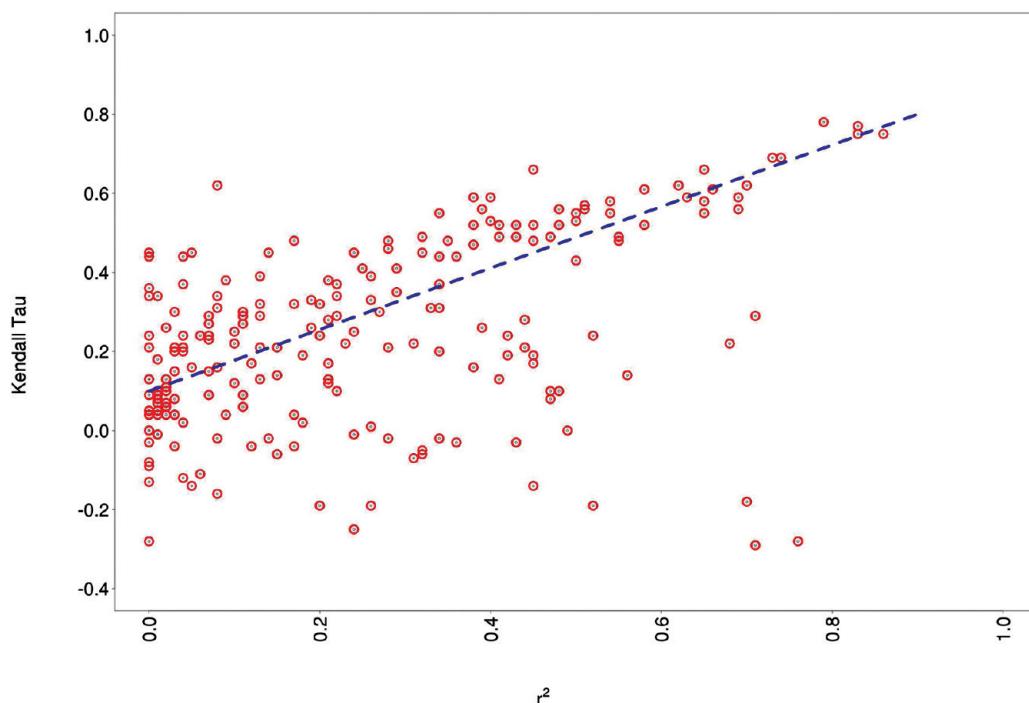
The 151 models created by these training sets are applied to the test set obtained during the first splitting of the original data. This provides a consistent performance reference for the model. In ROE evaluation, the modeling is performed using the Rensselaer Exploratory Center for Cheminformatics Research Online Modeling System or ROMS. ROMS is based on the Analyze PLS/KPLS software package version 6.96 (M. J. Embrechts, 2006). The functionalities of ROMS utilized by ROE evaluations include partial least-squares (PLS) and kernel partial least-squares (KPLS) regression modeling.^{7,8} Models created for ROE evaluation utilize 100 bootstraps, where the validation set for each bootstrap is ten percent of the training set size. To reduce computational time, parameter optimization was omitted for each truncation, and a set of standard model parameter values was used throughout. In this case, the standard values were 5 latent variables and 85% cousin threshold for both PLS and KPLS modeling and a sigma value of 10 for KPLS kernels.

Table 1. Data Sets Used in ROE Evaluation

data set	shorthand	num. mols.	target
angeotensin-converting enzyme	ACE	114	high blood pressure ¹³
acetylcholinesterase	AChE	60	alzheimer's/dementia ¹⁴
arteminisin analogs	art	179	malaria ¹⁵
boiling point	bp	298	¹⁶
HIV-RT	hivrt	64	HIV ¹⁷
Lombardo	lombardo	70	many ¹⁸
BBB	bbb	60	many ¹⁹
Benigni (mouse)	benigni.mouse	316	predicting toxicity ²⁰
Benigni (rat)	benigni.rat	375	predicting toxicity ²⁰
Comsia	comsia	88	blood clotting ²¹
COX-2	cox2	305	pain relief ²²
melting point	mp	277	²³
solubility	sol	1144	²⁴
steroids	steroids	31	depression ²⁵
Caco-2	caco2	45	intestinal transport ²⁶
ether-a-go-go	ether	126	long QT syndrome ²⁷
intestinal absorption	intabs	100	intestinal transport ²⁸
jejunum permeability	jperm	22	intestinal transport ²⁹
minnow toxicity	mintonx	322	toxicity ³⁰
oral absorption	oabs	23	intestinal transport ³¹
oral bioavailability	obio	275	intestinal transport ³²
P-glycoprotein	pgly	113	intestinal transport ³³
plasma binding	plasma	273	effective blood conc ³⁴
volume distribution.fu	lom.fu	120	effective blood conc ³⁵
volume distribution.vd	lom.vd	120	effective blood conc ³⁵
hERG	herg	86	long QT syndrome ³⁶
hERG 2	herg2	34	long QT syndrome ³⁷
hERG 3	herg3	65	long QT syndrome ³⁸
hERG 4	herg4	22	long QT syndrome ³⁹
hERG 5	herg5	67	long QT syndrome ⁴⁰
hERG 6	herg6	31	long QT syndrome ⁴¹
hERG 7	herg7	76	long QT syndrome ⁴²
hERG 8	herg8	101	long QT syndrome ⁴³
glucocorticoid receptor	gluc	35	inflammatory/autoimmune ^{44,45}
melanocortin-4 ag/antag	mel	82	obesity/cachexia ^{46,47}
PDGFR	pdgfr	77	cell proliferation ⁴⁸
5-lipoxygenase	slip	41	inflammation ⁴⁹
$\alpha 4\beta 2$ nicotinic acetylcholine receptor	a4b2	55	analgesic ⁵⁰
acid blockers	acid	38	stomach acid ⁵¹
adenosine A1 receptor	ad	32	heartrate ⁵²
nicotine acetylcholine agonist	agac	57	analgesic ⁵³
androgen	androgen	24	hormone replacement ⁵⁴
bradykinin	brady	34	chronic pain ⁵⁵
B-Raf	braf	37	cancer ⁵⁶
cannabinoid-1 receptor	cb1r	57	obesity ⁵⁷
cyclin-dependent kinase 4	cdk4	52	cancer ⁵⁸
carboxylesterase	cester	49	drug metabolism ⁵⁹
3-hydroxy 3-mythoglutaryl CoA reductase	coared	26	heart disease ⁶⁰
corticotropin-releasing factor receptor 1	crf1	44	depression/anxiety ⁶¹
cyanoguanidine P2X7	cyano	59	pain relief ⁶²
dehydrosqualine synthase	dhsqual	37	inhibit <i>Staphylococcus aureus</i> ⁶³
factor Xa	fxa	49	anticoagulant ¹⁷
glycerol 3-phosphate acyltransferase	g3pat	36	obesity ⁶⁴
hepatitis C virus NS3 serine protease	hepc	34	hepatitis ⁶⁵

Table 1. Continued

data set	shorthand	num. mols.	target
hepatitis C virus NS3 helicase	helicase	38	hepatitis ⁶⁶
histamine H3 receptor	hish3	35	CNS disorders ⁶⁷
HIV1 reverse transcriptase DABO analogues	hiv1	33 or 27	HIV ⁶⁸
interleukin 8	il8	35	inflammation ⁶⁹
ketopiperidinens histamine H3	ketopip	51	hay fever ⁷⁰
hormone-selective lipase inhibitors	hsl	30	insulin resistance ⁷¹
human monoamine oxidase A and B	mao	37 or 51	depression/oxidative stress ⁷²
mitogen-activated protein kinase kinase	mapkk	26	cancer ⁷³
MAPK-activated protein kinase 2	mkapk2	31	arthritis ⁷⁴
antimycobacterials	mycobac	35	tuberculosis ⁷⁵
Nek2/Hec1	nek2hec1	27	cancer ⁷⁶
procollagen C proteinase	pcp	37	topical antiscarring ⁷⁷
integrin $\alpha 2/\beta 1$	platad	26	cancer/clotting ⁷⁸
antiprotozoals	pzoan	37	African sleeping sickness ⁷⁹
stearyl-CoA desaturase 1	scd	48	obesity ⁸⁰
signal transducer/activation of transcription 3	stat3	49	cancer ⁸¹
vasopresin V2 receptor	vaso	45	water balance ⁸²

**Figure 3.** Correlation between Kendall Tau and r^2 for PLS models. The dashed blue line follows the rough correlation between r^2 and Kendall Tau.

The ROMS modeling process involved calculating r^2 , Q^2 , and RMSE metrics for each model. Calculation of rank order metrics was then performed across all 151 prediction results after ROMS modeling was complete, based on ranks derived from regression predictions. Kendall Tau for each model was calculated by comparing the predicted activity ranks to the actual activity ranks of the test set. The values of Kendall Tau for the ten models at each truncation stage are averaged, and then Shannon entropy was computed as a metric of Kendall Tau stability over the truncations.

Calculating 151 models for each data set, descriptor set, and modeling method combination is a computationally intensive process. During the investigation, it was observed that the behavior of a given data set, descriptor set, and modeling method combination could be assessed using only a portion of the 151 training sets. As shown in Figure 2, the original training set and the first four truncations convey the behavior of the majority of truncations. As the most change is often seen in the last few truncations, only the original training set, the first four truncations, and the last truncation are used in ROE evaluations to streamline the process.

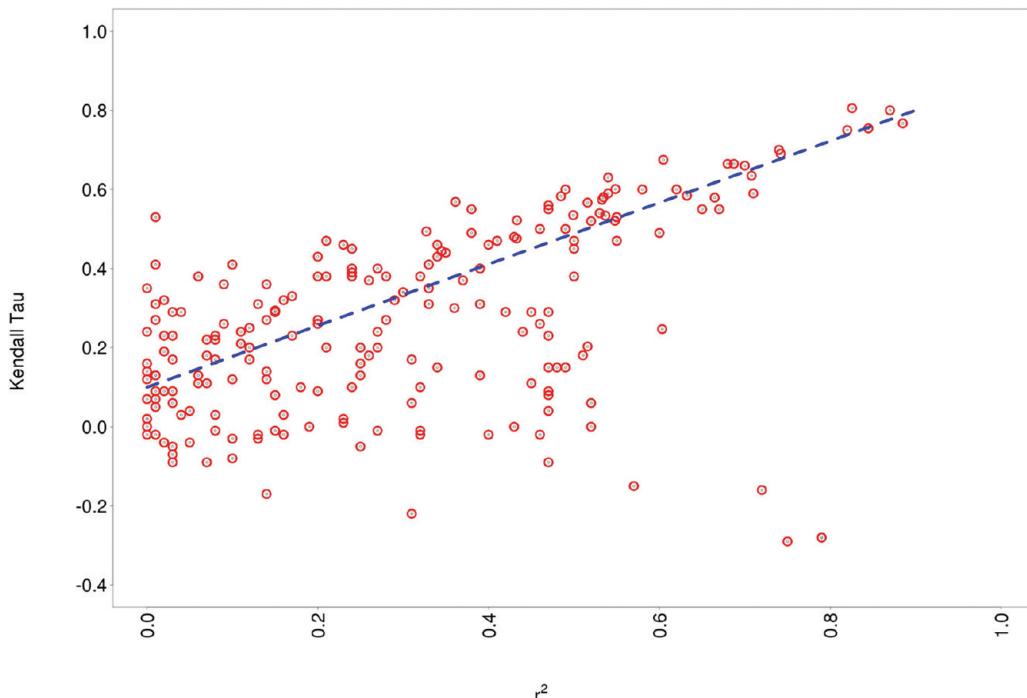


Figure 4. Correlation between Kendall Tau and r^2 for KPLS models. The dashed blue line follows the rough correlation between r^2 and Kendall Tau.

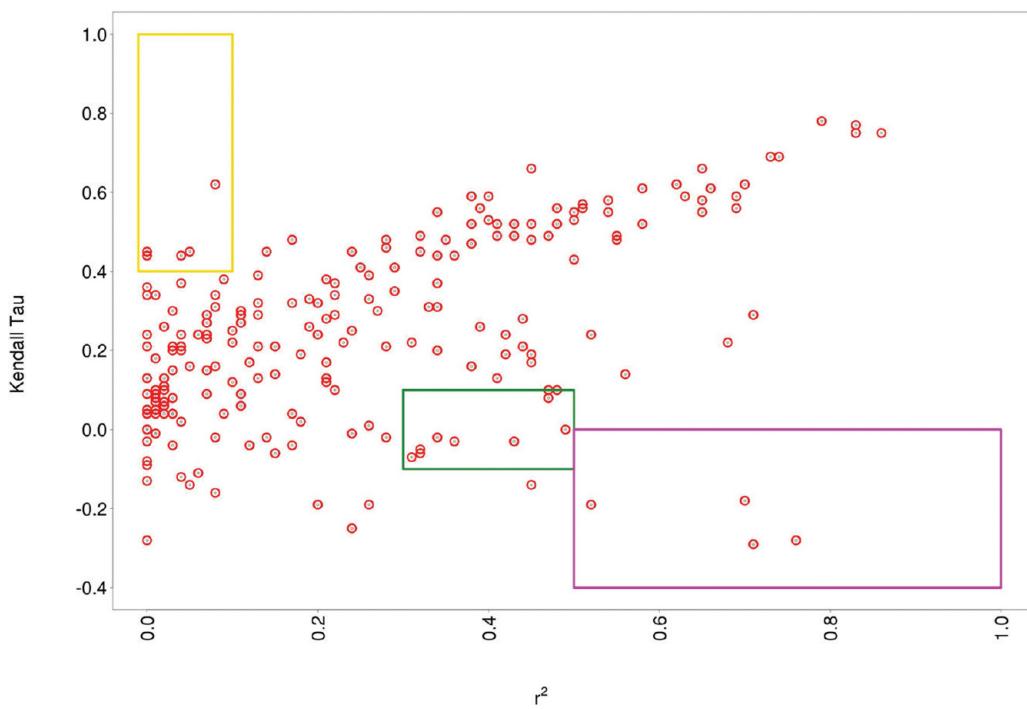


Figure 5. Comparing Kendall Tau and r^2 for PLS models. The boxes enclose subsets of outliers. The gold box encloses low r^2 /high Kendall Tau outliers which are often discarded. The green box encloses models on the threshold of acceptable r^2 values which are sometimes discarded. The pink box encloses models that are often used, potentially with poor results.

Large, negative Kendall Tau values were observed in initial tests of some combinations of data, descriptor, and modeling method. While such values of Kendall Tau usually indicate a negative correlation with true rank order, it is possible that some

data sets contain activity values with experimental errors large enough to cause confusion in true rank order. To accommodate this, an ε -insensitive modeling provides a dead band within response values so models do not end up representing noise.

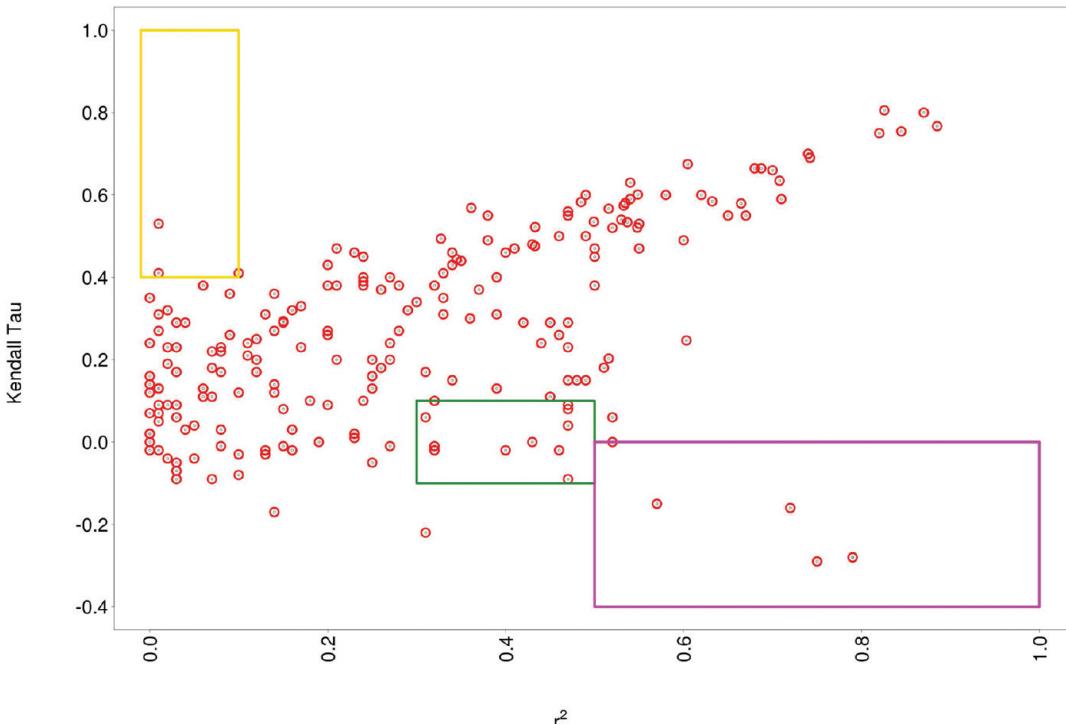


Figure 6. Comparing Kendall Tau and r^2 for KPLS models. The boxes enclose subsets of outliers according to the same color scheme as in Figure 5.

Table 2. PLS Data Sets of Interest

high r^2 and low KT	KT near zero	r^2 near zero
sol/moei3d ²⁴	cester/ra3d ⁵⁹	cox2/ra2d ²²
sol/moe2d ²⁴	hepc.ec90/moe2d ⁶⁵	cox2/ra3d ²²
steroids/moei3d ²⁵	hish3/TAE ⁶⁷	oabs/ra3d ³¹
steroids/moe2d ²⁵	jperm/moe2d ²⁹	herg4/moe2d ³⁹
	jperm/moei3d ²⁹	ache/TAE ¹⁴
	lom.vd/moei3d ³⁵	
	nek2hec1/TAE ⁷⁶	
	pzoan/moe2d ⁷⁹	
	pzoan/ra3d ⁷⁹	

Kendall Tau can be applied in an ε -insensitive way to explore the behavior of some of the outlying models using a sensitivity test, where the value of ε was varied based on the range and value of the activities of the molecules predicted by the model as well as an estimate of the experimental error.

The data sets used to test the ROE metric were selected from QSAR data sets in literature, curated online data sets, and industry databases. The 71 data sets used in the present study are shown in Table 1 and vary widely in size and type of target. Preparing the data sets for modeling and evaluation involved preprocessing molecular structures and setting reasonable ionization states as well as performing descriptor calculations. Both the preprocessing and descriptor calculation steps were performed using MOE version 2008.10 (Molecular Operating Environment, Chemical Computing Group Inc., Montreal, Canada). Six sets of descriptors were calculated for each data set, including MOE 2D (moe2d) descriptors, MOE orientation-independent 3D (moei3d) descriptors, reconstructed electron charge

Table 3. KPLS Data Sets of Interest

high r^2 and low KT	KT near zero	r^2 near zero
sol/moei3d ²⁴	g3pat/TAE ⁶⁴	bbb/moei3d ¹⁹
sol/tae ²⁴	g3pat/ra3d ⁶⁴	hish3/moe2d ⁶⁷
sol/moe2d ²⁴	dhsqual/ra3d ⁶³	oabs/ra3d ³¹
steroids/moei3d ²⁵	cester/ra3d ⁵⁹	herg4/TAE ³⁹
	g3pat/ra2d ⁶⁴	
	pzoan/ra3d ⁷⁹	
	hepc.ec90/moe2d ⁶⁵	
	cester/TAE ⁵⁹	
	mp/moe2d ²³	
	lom.vd/moei3d ³⁵	
	hish3/moei3d ⁶⁷	

densities and electron density-derived (RECON) 2-dimensional autocorrelated (ra2d) descriptors, RECON 3-dimensional autocorrelated (ra3d) descriptors, Transferable Atom Equivalent (TAE) descriptors, and ultrafast shape recognition (USR) descriptors.^{9–12} Each combination of data, descriptor, and modeling method is considered separately in the examination of the ROE evaluation process.

RESULTS AND DISCUSSION

Comparing rank order and traditional metrics on all 71 data sets yielded the results shown in Figures 3 and 4. Most data set, descriptor set, and modeling method combinations show a rough correlation between Kendall Tau and r^2 , as expected, indicated by the dashed red trend line in Figures 3 and 4. However, a handful of combinations behaved differently than expected. Figures 5 and 6 show the behavior of many combinations, including those that

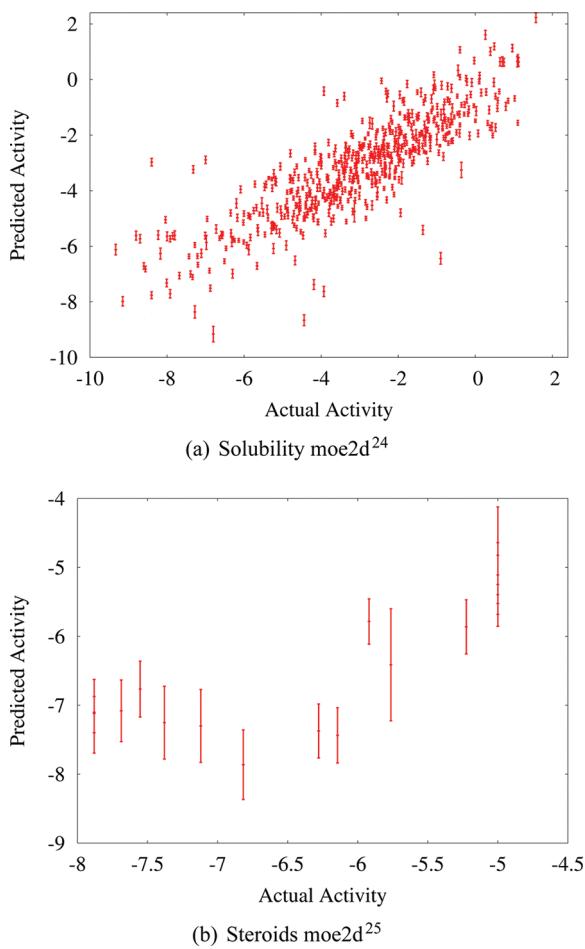


Figure 7. High r^2 PLS outliers solubility/moe2d and steroids/moe2d.

have unexpected values of Kendall Tau boxed in pink, green, or gold.

Of particular significance are the ROE evaluations boxed in pink; these have high r^2 values and very low Kendall Tau values. High r^2 values indicate high correlation between predicted and actual activities, and low Kendall Tau values indicate that predicted rank order does not correlate with actual rank order. Unfortunately, these models are often used when they should be discarded, or at minimum, examined in greater detail. The evaluations boxed in green have r^2 values greater than 0.6 and Kendall Tau values that indicate a lack of predictive ability for rank order. In many design applications, models with r^2 values greater than 0.6 are considered acceptable or "predictive" models. The third set of evaluations, boxed in gold, have r^2 values at or near zero, implying little or no correlation between predicted and actual activity values but still have high Kendall Tau values. These gold boxes contain models that would usually be discarded based on r^2 values but are examined here in more detail due to high Kendall Tau values.

The data sets used to create the models that fall within the outlying ranges are shown in Table 2 and Table 3. It was expected that different data sets might appear in each of the outlying sections when PLS and KPLS regression models were compared, since it was considered possible that the KPLS models would capture nonlinear relationships between descriptors and activity that were not captured in PLS models. However,

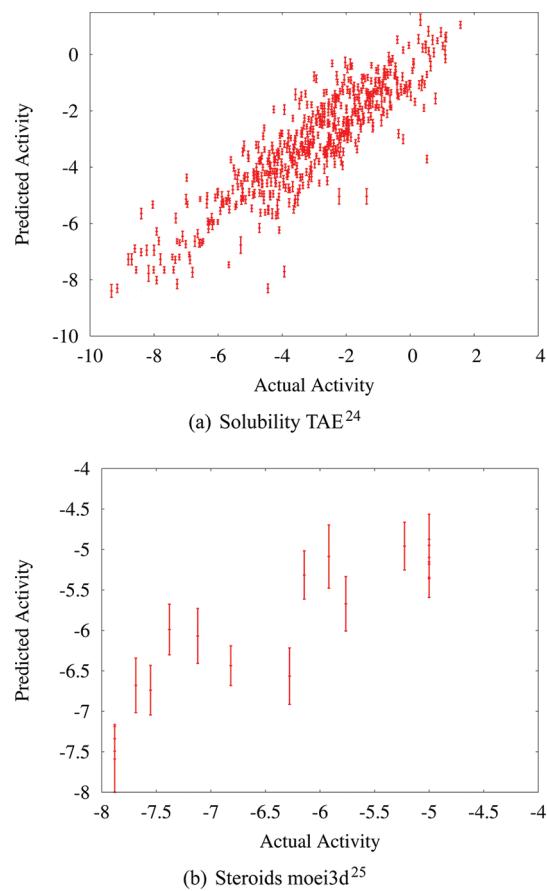


Figure 8. High r^2 KPLS outliers solubility/TAE and steroids/moei3d.

when tested, PLS and KPLS modeling produced similar results in Kendall Tau over all data sets. This suggests that any nonlinear relationships within these data sets were not particularly significant. Examples from each set of outliers for each modeling method were then examined in greater detail in an attempt to understand the evaluations, including analysis of the modeling output through the comparison of actual versus predicted activity.

In the PLS modeling portion of the ROE evaluation, the pink outliers with high r^2 and negative Kendall Tau values included a solubility data set using moe2d descriptors and a steroid data set using moe2d descriptors.^{24,25} The solubility data set is very large, including over 1100 molecules. Using traditional wisdom, it might be assumed that models created using large data sets with high r^2 values should be highly predictive, though this is not always the case. Figure 7 demonstrates that the rank orders of the molecules in this data set are not predicted well even though a clear regression line exists. In fact, the Kendall Tau value for this particular evaluation was below -0.2 , indicating a net reversal of order. Perhaps this is evidence that no more than a classifier model should be used for data sets of this type.

Though the steroid data set is much smaller than the solubility data set, this same problem is visible in terms of vastly different predicted rank order with some reversal of segments of rank ordering. The steroid data set with moe2d descriptors shown in Figure 7 has a second problem: the range of predicted values for each of the molecules varies a great deal over different bootstraps.

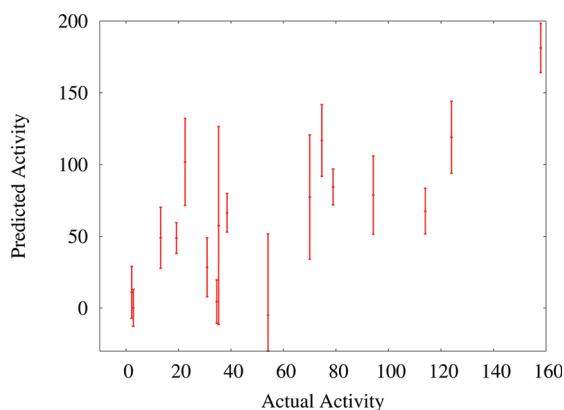
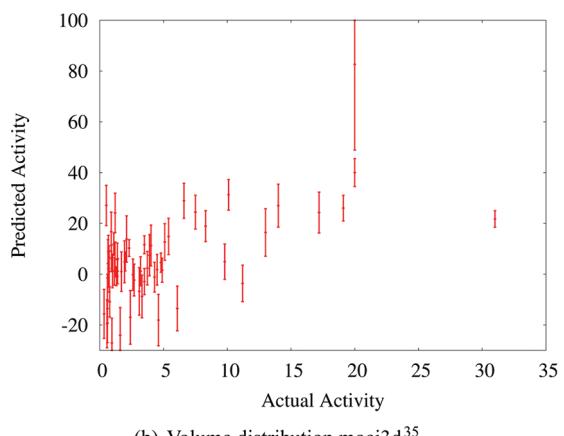
(a) Antiprotozoals moe2d⁷⁹(b) Volume distribution moe3d³⁵

Figure 9. Kendall Tau near zero PLS outliers antiprotozoals/moe2d and volume distribution/moei3d.

This may be due to the small size of the data set, as it contains only 36 molecules.²⁵ Future analysis could include evaluation of the distribution of molecular clusters across the training and testing sets to ascertain if the distribution of molecules has more of an effect than the information content of the descriptors.

The same range of predicted values is seen in the KPLS model outliers: the solubility data set with TAE descriptors and the steroid data set with moe2d descriptors as shown in Figure 8. Though the TAE descriptors have a strong monotonic relationship with molecular solubility in terms of r^2 , the rank order prediction using these descriptors is surprisingly ineffective. Regardless of data set size or the use of linear or nonlinear modeling, these data sets exhibit negative Kendall Tau values. If these four apparently good outlying models were used to predict the behavior of molecules with unknown activities, the rank order predictions would be unexpectedly poor.

The second set of outliers includes models with r^2 values approaching traditionally acceptable levels, but with Kendall Tau values near zero. Shown in Figure 9 are two PLS model outliers: the antiprotozoals data set using moe2d descriptors and the volume distribution data set using moe3d descriptors.^{35,79} The antiprotozoal/moe2d modeling data shows both complete reordering of predicted activity as well as varying sizes of prediction error bars. Further examination of the behavior of the model data shows a negative Q^2 value, indicating a high degree of overfitting. Both models shown in Figure 9 show negative Q^2 values. Though

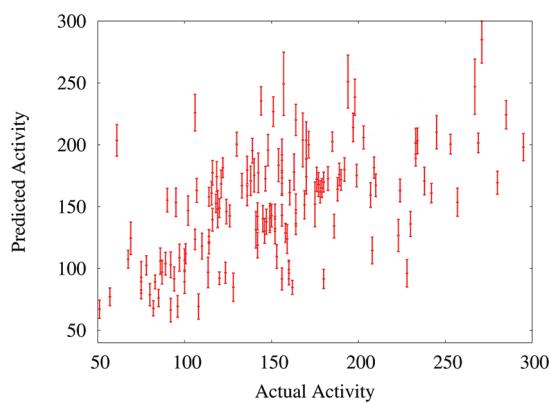
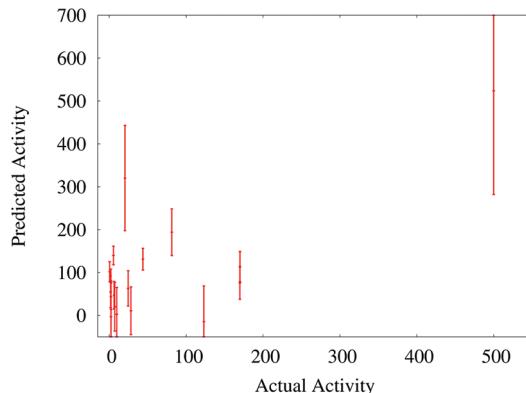
(a) Melting Point moe2d²³(b) Dehydrosqualine ra3d⁶³

Figure 10. Kendall Tau near zero KPLS outliers melting point/moe2d and dehydrosqualine/ra3d.

the volume distribution/moei3d model contains more molecules than the antiprotozoal/moe2d model, the volume distribution/moei3d model predictions incorporate more variance, as shown by the larger negative Q^2 value. Again, no sections of rank order are retained within the predicted activity values from this model.

The KPLS model outliers for r^2 values between 0.3 and 0.5 and Kendall Tau values near zero are shown in Figure 10. These include the melting point data set using moe2d descriptors and the dehydrosqualine data set using ra3d descriptors.^{23,63} The melting point/moe2d model contains more molecules than the dehydrosqualine/ra3d model, but only the dehydrosqualine/ra3d model exhibits a negative Q^2 value. Though the Q^2 value of the melting point/moe2d model is positive, it is tiny. The differing Q^2 performances could be due to the smaller range of scatter on the melting point/moe2d model data shown in Figure 10. Neither model retains segments of rank order, unlike the first set of outliers. These outliers demonstrate the necessity of using additional metrics when examining models with r^2 approaching acceptable values.

The third set of outliers include models with r^2 values near zero and large, positive Kendall Tau values. Figure 11 shows two such PLS model outliers: the COX-2 data set using ra2d descriptors and the acetylcholinesterase data set using TAE descriptors.^{14,22} Q^2 values for each of these models are well below zero, indicating that the models are highly overfitted. The graphs show that segments of rank order are retained within the

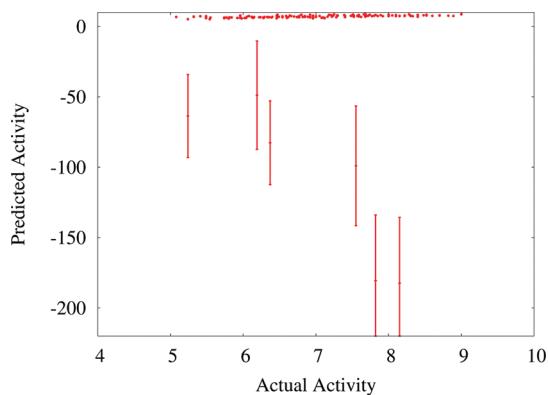
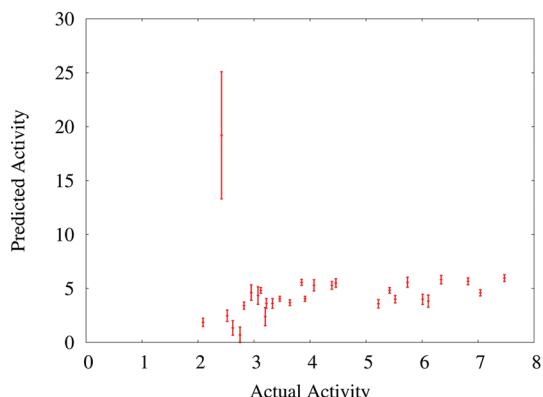
(a) COX-2 ra2d²²(b) Acetylcholinesterase TAE¹⁴

Figure 11. r^2 near zero PLS outliers COX-2/ra2d and acetylcholinesterase/TAE.

activity predictions, despite the lack of linear correlation between predicted and actual molecular activities. Though the KPLS models shown in Figure 12 contain fewer molecules than the PLS models, it is clear that the hERG 4 data set using TAE descriptors and the oral absorption data set using ra3d descriptors models have the same problems.^{31,39} The Q^2 values for both KPLS models are negative, though not as large as the Q^2 values for the PLS models. The small subsets of rank order that are retained within these KPLS models are also visible. While it is tempting to utilize the rank predictive power present in these models, the poor r^2 and Q^2 performance emphasize the necessity of utilizing more than one metric. Models with r^2 values near zero and large, positive Kendall Tau values should not be utilized in prediction.

Due to the number of ROE evaluations that showed negative Kendall Tau values, ε -insensitive Kendall Tau was chosen to examine five ROE evaluations. Applying this variant of Kendall Tau as a sensitivity test to the ROE evaluations involved examining the ranges of molecular activities and the size of expected experimental error for each of the data sets to determine the range of values of ε to apply to each ROE evaluation. ε -insensitive Kendall Tau was used to examine representatives of all three sets of outliers as well as evaluations with Kendall Tau values that correlated with r^2 values. All five evaluations were chosen based on data set size, with a preference for larger data sets that had a greater probability of containing molecules with similar activities. The hypothesis was that ε -insensitive Kendall

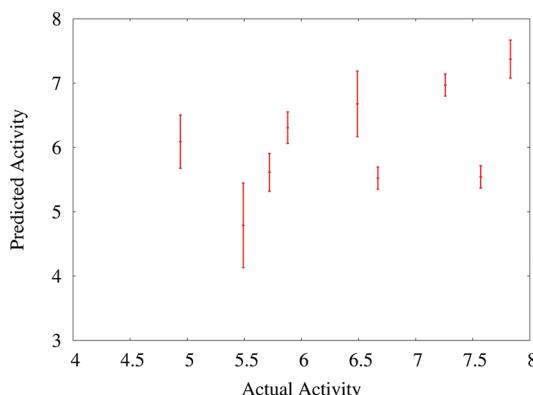
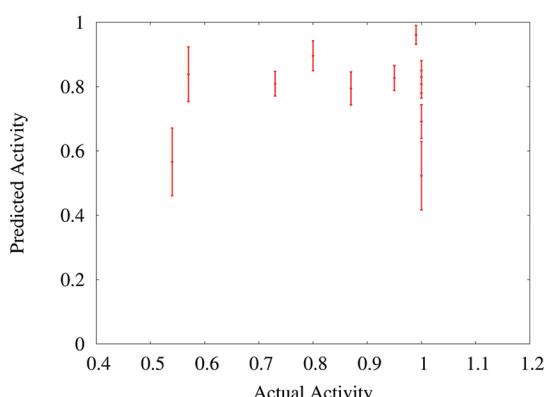
(a) hERG 4 TAE³⁹(b) Oral absorption ra3d³¹

Figure 12. r^2 near zero PLS outliers hERG 4 TAE and oral absorption ra3d.

Tau results would reveal whether experimental error had a noticeable effect on Kendall Tau performance.

The ROE outlier evaluations selected for ε -insensitive testing were as follows: 1) the solubility set using moe2d descriptors, 2) the volume distribution set using moei3d descriptors, and 2) the COX-2 set using ra2d descriptors. In Figure 13, Figure 14, and Figure 15, the sensitivity tests of ε -insensitive Kendall Tau show minimum change over varying ε values. The nonoutlier evaluations selected were the ACE inhibitors set using moe2d descriptors and the boiling point set using TAE descriptors. Figures 16 and 17 show more change in Kendall Tau over ε values than the outlier ε -insensitive Kendall Tau sensitivity tests. The difference in behavior indicates that experimental error in data collection did not cause the outlying values of Kendall Tau.

Since one of the goals of the ROE project was to determine the relationship between ranking accuracy and the stability of rank order, evaluation of ROE required the definition of a suitable stability metric, involving the behavior of Kendall Tau over multiple data set truncations. Stability was measured using the Shannon entropy of Kendall Tau over the truncations.

Examining the relationship between rank order stability and rank order performance is possible when Shannon entropy is plotted against Kendall Tau. Figures 18 and 19 comparing Shannon entropy and Kendall Tau show a lack of correlation for both PLS and KPLS models. This lack of a relationship between stability and rank order emphasizes the value of ROE evaluation as an independent means for assessing the behavior of

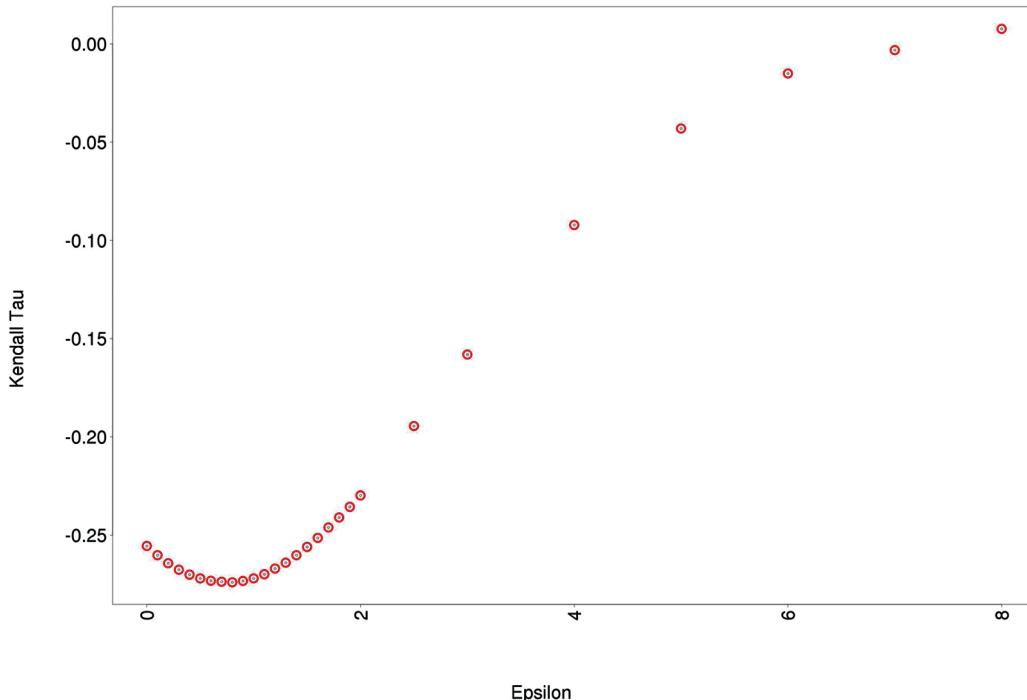


Figure 13. Exploring sensitivity of ε in ε -insensitive Kendall Tau: solubility Moe2d data.

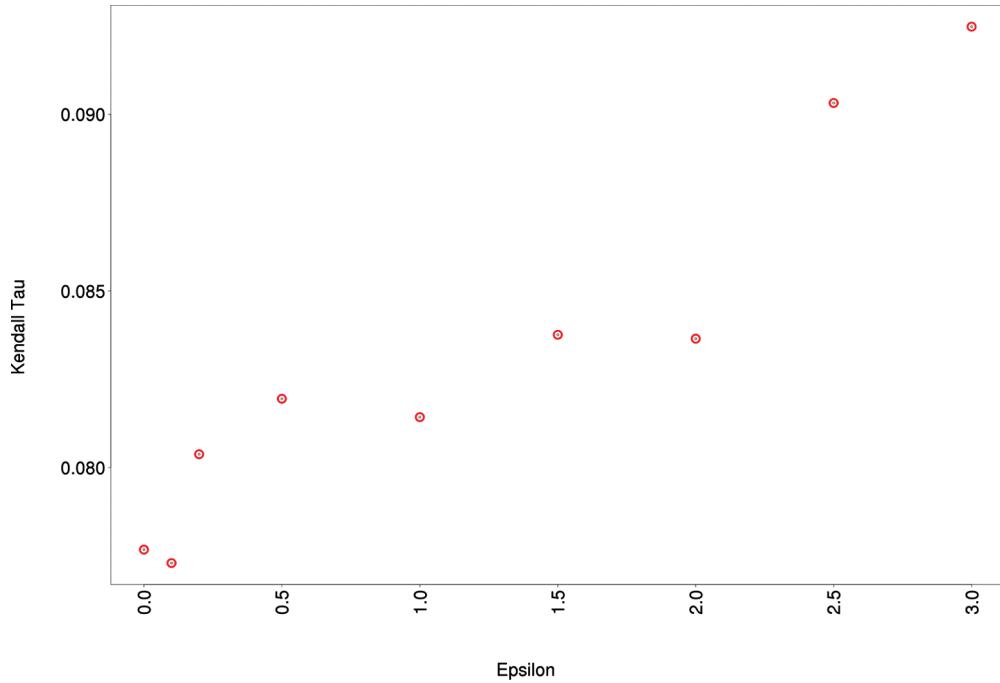


Figure 14. Exploring sensitivity of ε in ε -insensitive Kendall Tau: the volume distribution moei3d data.

a data set, descriptor set, and modeling method in prioritizing cases.

In practice, ROE evaluation requires a defined threshold that distinguishes between stable and unstable evaluations. To discern what value would be appropriate, Figure 20 illustrates the more stable behaviors of Shannon entropy of Kendall Tau for sets of six

Kendall Tau values. Each of these Kendall Tau values represents the evaluation of a combination of data, descriptors, and modeling method at a different truncation. Based on the values in the figure, only evaluations with a very small amount of change within Kendall Tau values would be classified as stable combinations of data set, descriptor set, and modeling method. For this reason, the threshold

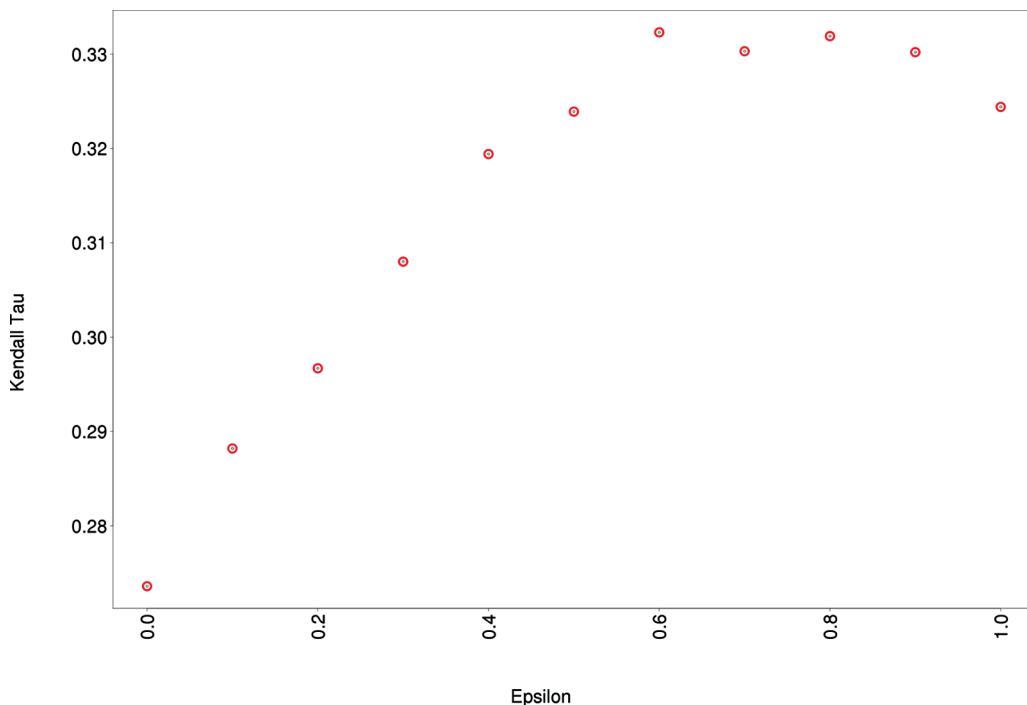


Figure 15. Exploring sensitivity of ε in ε -insensitive Kendall Tau: COX-2 ra2d data.

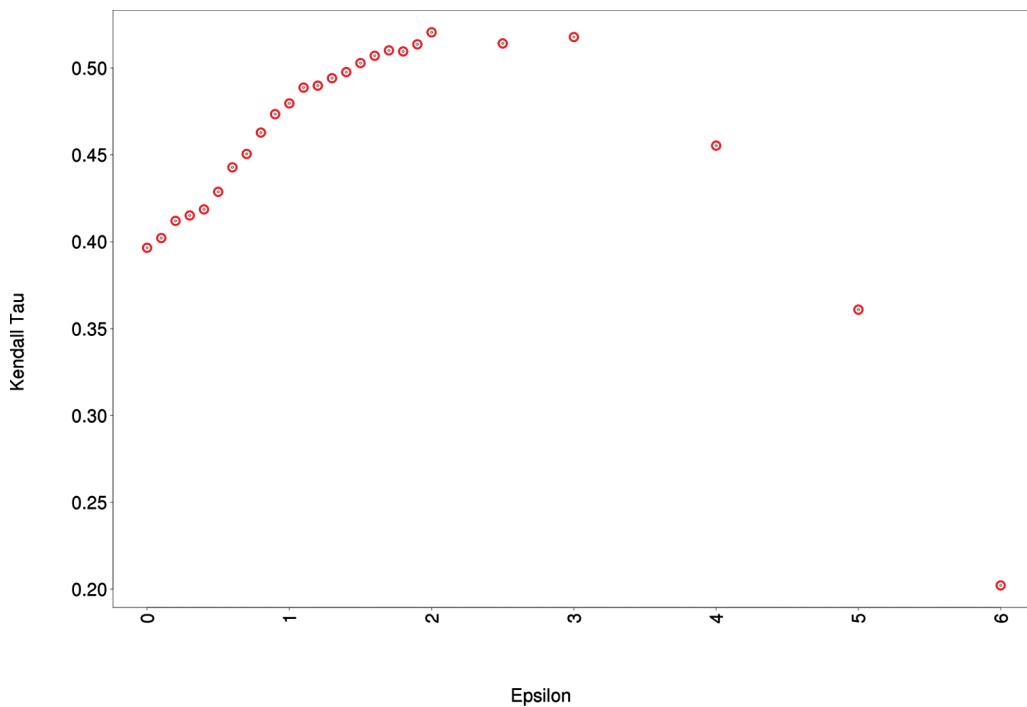


Figure 16. Exploring sensitivity of ε in ε -insensitive Kendall Tau: ACE Moe2d data.

for stability was set at 0.2, with stable evaluations falling below and unstable evaluations falling above this value.

The many combinations of data set, descriptor set, and modeling methods were classified using Kendall Tau and Shannon entropy of Kendall Tau, resulting in the contents of Table 4.

The behavior of the combinations in each of the classifications was examined and yielded the results shown in Table 5. Regardless of Kendall Tau values, if r^2 values for models do not meet minimum acceptable values, the model should be discarded. Due to the similarity in performance of combinations of data sets and

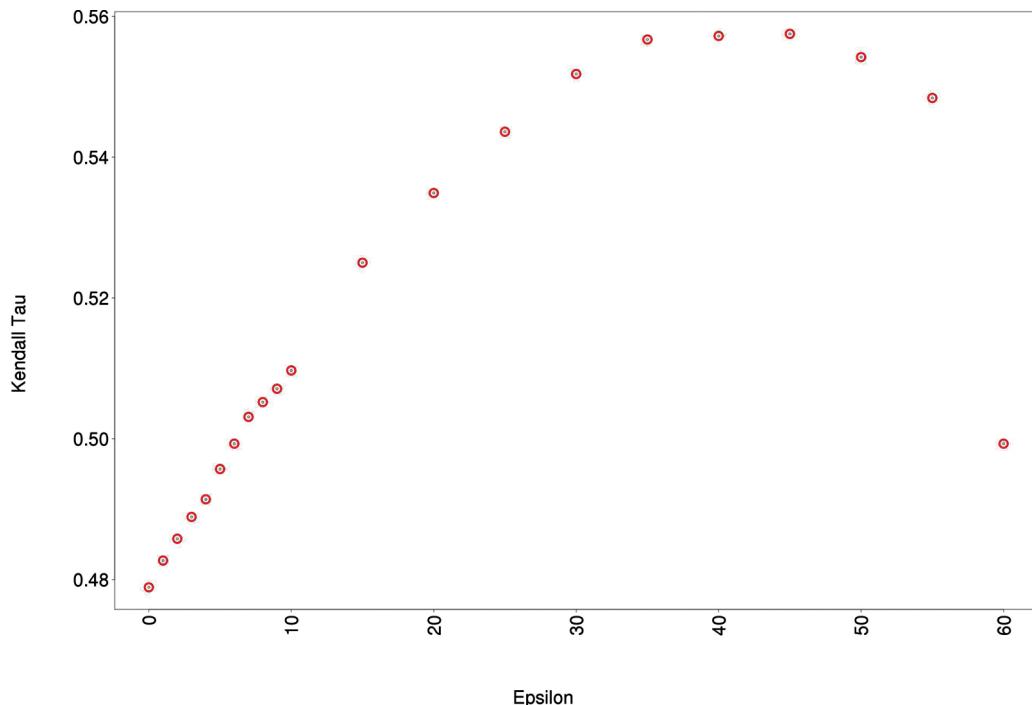


Figure 17. Exploring sensitivity of ε in ε -insensitive Kendall Tau: BP TAE data.

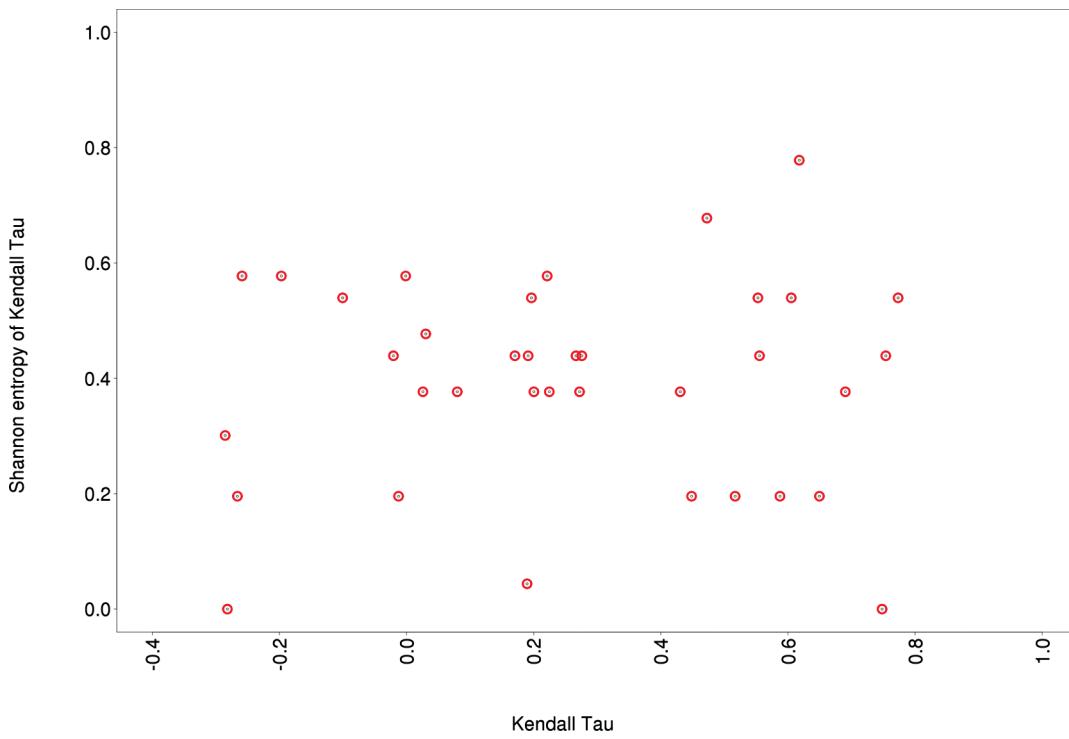


Figure 18. Shannon entropy as a measure of Rank Order Entropy: Kendall Tau data vs Shannon entropy for PLS models.

descriptors within PLS and KPLS, the recommendations are presented as a set of general guidelines, which are applicable to any modeling method.

Many of the recommendations in Table 5 were determined based on the performance of the models created using the data

sets listed in Table 4. There are some sections in Table 4 that are not populated, despite efforts to the contrary. These sections have no representative data within the combinations of data set, descriptor set, and modeling methods tested because combinations could not be found that performed in those ranges.

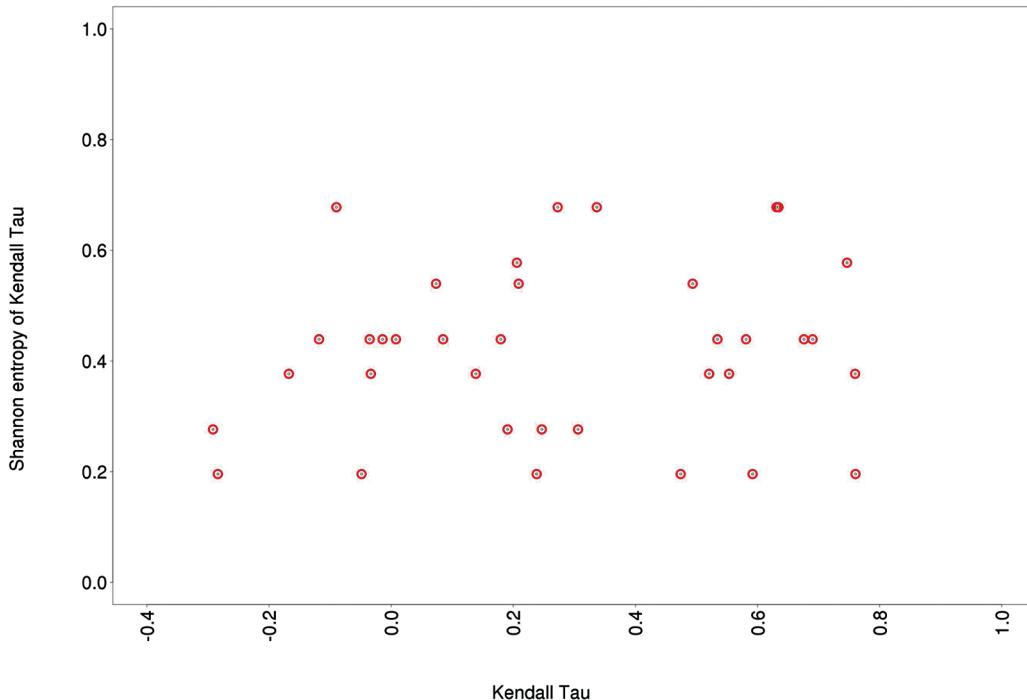


Figure 19. Shannon entropy as a measure of Rank Order Entropy: Kendall Tau data vs Shannon entropy for KPLS models.

Creating the recommendations in Table 5 began with assessing the ROE evaluations by the Kendall Tau value of the original training set. Since Kendall Tau values roughly correlate with r^2 values, and Kendall Tau values of 0.4 align with r^2 values near 0.6, a Kendall Tau value of 0.4 was set as the minimum acceptable value. Below that Kendall Tau value, rank order performance was observed to drop off. Above a Kendall Tau value of 0.6, performance improved significantly over models with Kendall Tau values between 0.4 and 0.6. A similar improvement in model performance was observed when Kendall Tau increased from 0.6 to 0.8. When Kendall Tau values fell between 0.2 and 0.4, model performance diminished. Examining ROE results led to the recommendation that an addition of data to the model could improve models that had Kendall Tau values between 0.2 and 0.4, though a change in descriptors would potentially aid model performance as well. The range of Kendall Tau from -0.2 to 0.2 appeared to contain models devoid of information. Whether this lack of information was due to lack of representation of appropriate chemical properties by the chosen descriptors or simply an inadequate number of cases within the data set likely varies from combination to combination. ROE evaluations with Kendall Tau values below -0.2 clearly need a change in the model creation process due to the reversal of rank order prediction. It is likely that these combinations require either a new descriptor set or a new modeling method.

When stability of Kendall Tau is added to the evaluation, the assessments for combinations of data set, descriptor set, and modeling method become more specific. Stable combinations receive assessments based on the Kendall Tau value of the original training set, as those values change minimally over the truncations. Unstable combinations require more cautious recommendations. As stated earlier, the utility of a model is directly related to the stability of rank order prediction. The most essential recommendation provided for unstable combinations is to obtain more information, either by changing descriptor sets or by adding molecules to the data set. For combinations with large data sets, the data set size is likely not the problem, making

Combination	Truncations colored by bins					Shannon entropy
BP/moe3d						0.0
HIVRT/ra2d						0.1957
intabs/TAE						0.2764
sol/moe2d						0.3010
ACE/USR						0.3768

Figure 20. Shannon entropy values for various groupings of data. This table shows the possible combinations of different bins at the stable end of the spectrum. As can be seen, if five Kendall Tau values fall in one bin, with only one Kendall Tau value in a different bin, the Shannon entropy of the truncations is the smallest nonzero possibilities shown. A small value of Shannon entropy indicates a higher stability in the modeling process. The next largest Shannon entropy value comes from the combination of 2 Kendall Tau values in one bin, and 4 Kendall Tau values in another bin. The third largest Shannon entropy value comes from Kendall Tau values split evenly into two bins. The last example shown here is the fourth highest value of Shannon entropy, caused by one Kendall Tau value in 2 different bins, and the other 4 Kendall Tau values in one bin.^{13,16,24,28,83}

the descriptor set selection or increasing data set diversity more relevant recommendations. Though the performance of combinations with large data sets was not improved by applying an ϵ -insensitive Kendall Tau sensitivity test, combinations with small data sets may show an improvement in performance. These results suggest that quantitative analysis of rank order stability improves the confidence of using a particular combination of data, descriptors, and modeling methods in prioritization schemes.

Table 4. PLS and KPLS Data Classified Using Kendall Tau and Shannon Entropy

KT range	PLS data sets		KPLS data sets	
	stable	unstable	stable	unstable
0.6 to 0.8	bp/moei3d ¹⁶ hivrt/ra2d ¹⁷	ketopip/ra3d ⁷⁰ bp/tae ¹⁶ hivrt/ra3d ¹⁷ herg4/moe2d ³⁹ ace/moe2d ¹³	bp/moei3d ¹⁶	bp/tae ¹⁶ ketopip/ra3d ⁷⁰ hivrt/ra2d ¹⁷ herg4/moe2d ³⁹
0.4 to 0.6	mintonx/moe2d ³⁰ ace/usr ¹³ cox2/ra2d ²²	mintonx/tae ³⁰ hivrt/moei3d ¹⁷ lombardo/tae ¹⁸ ache/tae ¹⁴	mintonx/moe2d ³⁰ cox2/ra2d ²²	ache/tae ¹⁴ mintonx/tae ³⁰ ace/moe2d ¹³ ace/usr ¹³ lombardo/tae ¹⁸
0.2 to 0.4		comsia/ra3d ²¹ hish3/ra2d ⁶⁷ intabs/tae ²⁸ art/tae ¹⁵ Slip/moe2d ⁴⁹ art/moe2d ¹⁵ oabs/moe2d ³¹ il8/moei3d ⁶⁹ dhsqual/ra2d ⁶³ pgly/tae ³³ pzoan/ra3d ⁷⁹ braf/moe2d ⁵⁶ nek2hec1/ra3d ⁷⁶	comsia/ra3d ²¹	Slip/moe2d ⁴⁹ intabs/tae ²⁸ oabs/moe2d ³¹ art/tae ¹⁵ hish3/ra2d ⁶⁷ il8/moei3d ⁶⁹ art/moe2d ¹⁵ pgly/tae ³³ dhsqual/ra2d ⁶³ braf/moe2d ⁵⁶ pcp/moe2d ⁷⁷ pzoan/ra3d ⁷⁹
0.0 to 0.2		lom.vd/moei3d ³⁵	caco2/moei3d ²⁶	lom.vd/moei3d ³⁵ fxa/tae ¹⁷ g3pat/moe2d ⁶⁴ nek2hec1/ra3d ⁷⁶ steroids/moei3d ²⁵ g3pat/ra3d ⁶⁴
-0.2 to 0.0		g3pat/ra3d ⁶⁴ sol/tae ²⁴	steroids/moei3d ²⁵ sol/moe2d ²⁴	sol/tae ²⁴ sol/moe2d ²⁴

Table 5. Recommendations for Data Sets Based on KT and Shannon Entropy Performance

grade	KT range	stable	unstable
A	0.8 to 1.0	Use.	Use with caution.
B	0.6 to 0.8	Use.	Use with caution; ε -insensitive test recommended. If very unstable, needs new desc.
C	0.4 to 0.6	Use with caution. If possible, get more data/new desc.	Use with caution. ε -insensitive test recommended.
D	0.2 to 0.4	Do not use; needs more data.	Do not use. Needs new desc/more data.
F	-0.2 to 0.2	Do not use. Needs new desc/signal missing.	Do not use. Needs new desc/signal missing.
E	below -0.2	Do not use. Needs new descriptors/method.	Do not use. Needs new descriptors/method.

CONCLUSION

Development of the ROE evaluation method uncovered a relationship between rank order and test set r^2 performance as well as other unexpected information. Kendall Tau and r^2 were

found to be correlated but not in all cases. The outliers of this comparison have great significance, indicating that more caution is needed when using models with high r^2 values and that use of models with near-acceptable r^2 values require caution. For ROE

with large data sets, ε -insensitive Kendall Tau sensitivity tests reveal that the behavior of these models are not necessarily outliers because of experimental error. When rank order stability is compared with actual values of Kendall Tau for different models, the most striking observation is that no correlation exists between them. Stable ROE evaluations appear across the range of Kendall Tau, as do unstable ROE evaluations. ROE evaluations are designed to reveal the stability of rank order prediction and were tested using various combinations of data sets, descriptor sets, and modeling methods. This method enables intelligent use of QSAR models by increasing the confidence level of applying a given QSAR model to a particular problem by providing a metric that goes beyond traditional metrics of model quality assessment.

AUTHOR INFORMATION

Corresponding Author

*E-mail: brenec@rpi.edu.

ACKNOWLEDGMENT

RPI Center for Biotechnology and Interdisciplinary Studies
NIH Grant 1P20-HG003899-03.

REFERENCES

- (1) Bennett, K.; Campbell, C. Support Vector Machines: Hype or Hallelujah? *SIGKDD Explorations* **2000**, *2*, 1–13.
- (2) Boslaugh, S.; Watters, P. A. *Statistics in a Nutshell*; O'Reilly Media Inc: Sebastopol, CA, 2008; p 183.
- (3) Abdi, H. In *Encyclopedia of Measurement and Statistics*; Salkind, N., Ed.; Sage: Thousand Oaks, CA, 2007; Chapter The Kendall Rank Correlation Coefficient, pp 1–7.
- (4) Farrell, F. E. *Regression Modeling Strategies*, 2nd ed.; Springer, 2001; p 94, Accessed online Mar 12 2009.
- (5) Clark, R. D.; Fox, P. C. Statistical variation in progressive scrambling. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 563–576.
- (6) Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423 and 623–656.
- (7) Abdi, H. In *Encyclopedia of Social Sciences Research Methods*; Lewis-Beck, M., Bryman, A., Futing, T., Eds.; Sage: Thousand Oaks, CA, 2003; Chapter Partial Least Squares (PLS) Regression, pp 1–7.
- (8) Rosipal, R. Kernel Partial Least Squares for Nonlinear Regression and Discrimination. *Neural Network World* **2009**, *13*, 291–300.
- (9) Labute, P. A Widely Applicable Set Of Descriptors. *J. Mol. Graphics Modell.* **2000**, *18*, 464–477.
- (10) Breneman, C. M.; Sundling, C. M.; Sukumar, N.; Shen, L.; Katt, W. P.; Embrechts, M. J. New developments in PEST shape/property hybrid descriptors. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 231–240.
- (11) Whitehead, C. E.; Breneman, C. M.; Sukumar, N.; Ryan, M. D. Transferable Atom Equivalent Multi-Centered Multipole Expansion Method. *J. Comput. Chem.* **2003**, *24*, 512–529.
- (12) Ballester, P. J.; Finn, P. W.; Richards, W. G. Ultrafast shape recognition: Evaluating a new ligand-based virtual screening technology. *J. Mol. Graphics Modell.* **2009**, *27*, 836–845.
- (13) Southerland, J. J.; O'Brien, L. A.; Weaver, D. F. A Comparison of Methods for Modeling Quantitative Structure-Activity Relationships. *J. Med. Chem.* **2004**, *47*, 5541–5554.
- (14) Cho, S. J.; Garsia, M. L. S.; Bier, J.; Tropsha, A. Structure-Based Alignment and Comparative Molecular Field Analysis of Acetylcholinesterase Inhibitors. *J. Med. Chem.* **1996**, *39*, 5064–5071.
- (15) Guha, R.; Jurs, P. C. Development of QSAR Models to Predict and Interpret the Biological Activity of Artemisinin Analogs. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1440–1449.
- (16) Hall, L. H.; Story, C. T. Boiling Point and Critical Temperature of a Heterogeneous Data Set: QSAR with Atom Type Electrotopological State Indices Using Artificial Neural Networks. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1004–1014.
- (17) Ye, B.; et al. Thiophene-Anthranilamides as Highly Potent and Orally Available Factor Xa Inhibitors. *J. Med. Chem.* **2007**, *50*, 2967–2980.
- (18) Lombardo, F.; Blake, J. F.; Curatolo, W. J. Computation of Brain-Blood Partitioning of Organic Solutes via Free Energy Calculations. *J. Med. Chem.* **1996**, *39*, 4750–4755.
- (19) Liu, R.; Sun, H.; So, S.-S. Development of Quantitative Structure-Property Relationship Models for Early ADME Evaluation in Drug Discovery. 2. Blood-Brain Barrier Penetration. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1623–1632.
- (20) Quantitative Structure-Activity Relationship (QSAR) models of mutagens and carcinogens; Benigni, R., Ed.; CRC Press: Boca Raton, 2003.
- (21) Böhm, M.; Stürzebecher, J.; Klebe, G. Three-Dimensional Quantitative Structure-Activity Relationship Analyses Using Comparative Molecular Similarity Indices Analysis To Elucidate Selectivity Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa. *J. Med. Chem.* **1999**, *42*, 458–477.
- (22) Chavatte, P.; Yous, S.; Marot, C.; Baurin, N.; Lesieur, D. Three-Dimensional Quantitative Structure-Activity Relationships of Cyclooxygenase-2 (COX-2) Inhibitors: A Comparative Molecular Field Analysis. *J. Med. Chem.* **2001**, *44*, 3223–3230.
- (23) Bergström, C. A. S.; Norinder, U.; Luthman, K.; Artursson, P. Molecular Descriptors Influencing Melting Point and Their Role in Classification of Solid Drugs. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1177–1185.
- (24) Delaney, J. S. ESOL: estimating aqueous solubility directly from molecular structure. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1000–1005.
- (25) Coats, E. A. The CoMFA Steroids as a Benchmark Dataset for Development of 3D QSAR Methods. *Perspect. Drug Discovery Des.* **1998**, *12/13/14*, 199–213.
- (26) Nordqvist, A.; Nilsson, J.; Lindmark, T.; Eriksson, A.; Garberg, P.; Kihlén, M. A General Model for Prediction of Caco-2 Cell Permeability. *QSAR Comb. Sci.* **2004**, *23*, 303–310.
- (27) Ekins, S.; Balakin, K. V.; Savchuk, N.; Ivanenkov, Y. Insights for Human Ether-a-Go-Go-Related Gene Potassium Channel Inhibition Using Recursive Partitioning and Kohonen and Sammon Mapping Techniques. *J. Med. Chem.* **2006**, *49*, 5059–5071.
- (28) Raevsky, O. A.; Schaper, K.-J.; Artursson, P.; McFarland, J. W. A Novel Approach for Prediction of Intestinal Absorption of Drugs in humans based on Hydrogen Bond Descriptors and Structural Similarity. *Quant. Struct.-Act. Relat.* **2002**, *20*, 402–413.
- (29) Winiwarter, S.; Bonham, N. M.; Ax, F.; Hallberg, A.; Lennernäs, H.; Karlén, A. Correlation of Human Jejunal Permeability (in Vivo) of Drugs with Experimentally and Theoretically Derived Parameters. A Multivariate Data Analysis Approach. *J. Med. Chem.* **1998**, *41*, 4939–4949.
- (30) He, L.; Jurs, P. C. Assessing the reliability of a QSAR model's predictions. *J. Mol. Graphics Modell.* **2005**, *23*, 503–523.
- (31) Linnankoski, J.; Mäkelä, J. M.; Ranta, V.-P.; Urtti, A.; Yliperttula, M. Computational Prediction of Oral Drug Absorption Based on Absorption Rate Constants in Humans. *J. Med. Chem.* **2006**, *49*, 3674–3681.
- (32) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623.
- (33) Varma, M. V. S.; Sateesh, K.; Panchangnula, R. Functional Role of P-Glycoprotein in Limiting Intestinal Absorption of Drugs: Contribution of Passive Permeability to P-Glycoprotein Mediated Efflux aTransport. *Mol. Pharmaceutics* **2005**, *2*, 12–21.
- (34) Goodman and Gilman's *The Pharmacological Basis of Therapeutics*, tenth ed.; Hardman, J. G., Limbird, L. E., Gilman, A. G., Eds.; McGraw-Hill: 2001; pp 1917–2023.
- (35) Lombardo, F.; Obach, R. S.; Shalaeva, M. Y.; Gao, F. Prediction of Human Volume of Distribution Values for Neutral and Basic Drugs. 2. Extended Data Set and Leave-Class-Out Statistics. *J. Med. Chem.* **2004**, *47*, 1242–1250.
- (36) Bains, W.; Basman, A.; White, C. HERG binding specificity and binding site structure: evidence from a fragment-based evolutionary computing SAR study. *Prog. Biophys. Mol. Biol.* **2004**, *86*, 205–233.

- (37) Gavaghan, C. V.; Arnby, C. H.; Blomberg, N.; Strandlund, G.; Boyer, S. Development, interpretation and temporal evaluation of a global QSAR of hERG electrophysiology screening data. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 189–206.
- (38) Sun, H. An Accurate and Interpretable Bayesian Classification Model for Prediction of hERG Liability. *Chem. Med. Chem.* **2006**, *1*, 315–322.
- (39) Recanatini, M.; Poluzzi, E.; Masetti, M.; Cavalli, A.; Ponti, F. D. QT Prolongation Through hERG K⁺ Channel Blockade: Current Knowledge and Strategies for the Early Prediction During Drug Development. *Med. Res. Rev.* **2005**, *25*, 133–166.
- (40) Keseru, G. M. Prediction of hERG Potassium Channel Affinity by Traditional and Hologram QSAR Models. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2773–2775.
- (41) Cavalli, A.; Poluzzi, E.; Ponti, F. D.; Recanatini, M. Toward a Pharmacophore for Drugs Inducing the Long QT Syndrome: Insights from a CoMFA Study of hERG K⁺ Channel Blockers. *J. Med. Chem.* **2002**, *45*, 3844–3853.
- (42) Yoshida, K.; Niwa, T. Quantitative Structure-Activity Relationship Studies on Inhibition of hERG Potassium Channels. *J. Chem. Inf. Model.* **2006**, *46*, 1371–1378.
- (43) Song, M.; Clark, M. Development and Evaluation of an in Silico Model for hERG Binding. *J. Chem. Inf. Model.* **2006**, *46*, 392–400.
- (44) Takahashi, H.; Bekkali, Y.; Capolino, A. J.; Gilmore, T.; Goldrick, S. E.; Nelson, R. M.; Terenzio, D.; Wang, J.; Zuvela-Jelaska, L.; Proudfoot, J.; Nabozny, G.; Thomson, D. Discovery and SAR study of novel dihydroquinoline containing glucocorticoid receptor ligands. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1549–1552.
- (45) Takahashi, H.; Bekkali, Y.; Capolino, A. J.; Gilmore, T.; Goldrick, S. E.; Kaplita, P. V.; Liu, L.; Nelson, R. M.; Terenzio, D.; Wang, J.; Zuvela-Jelaska, L.; Proudfoot, J.; Nabozny, G.; Thomson, D. Discovery and SAR study of novel dihydroquinoline-containing glucocorticoid receptor agonists. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5091–5095.
- (46) Tran, J. A.; et al. Pyrrolidinones as orally bioavailable antagonists of the human melanocortin-4 receptor with anti-cachectic activity. *Bioorg. Med. Chem.* **2007**, *15*, 5166–5176.
- (47) Tran, J. A.; Chen, C. W.; Jiang, W.; Tucci, F. C.; Fleck, B. A.; Marinkovic, D.; Arellano, M.; Chen, C. Pyrrolidinones as potent functional agonists of the human melanocortin-4 receptor. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5165–5170.
- (48) Pandey, A.; Volkots, D. L.; Seroogy, J. M.; Rose, J. W.; Yu, J.-C.; Lambing, J. L.; Hutchaleelaha, A.; Hollenbach, S. J.; Abe, K.; Giese, N. A.; Scarborough, R. M. Identification of Orally Active, Potent, and Selective 4-Piperazinylquinazolines as Antagonists of the Platelet-Derived Growth Factor Receptor Tyrosine Kinase Family. *J. Med. Chem.* **2002**, *45*, 3772–3793.
- (49) Karg, E.-M.; Luderer, S.; Pergola, C.; Bühring, U.; Rossi, A.; Northoff, H.; Sautebin, L.; Troschütz, R.; Werz, O. Structural Optimization and Biological Evaluation of 2-Substituted 5-Hydroxyindole-3-carboxylates as Potential Inhibitors of Human 5-Lipoxygenase. *J. Med. Chem.* **2009**, *52*, 3474–3483.
- (50) Ji, J.; Schrimpf, M. R.; Sippy, K. B.; Bunnelle, W. H.; Li, T.; Anderson, D. J.; Faltynek, C.; Surowy, C. S.; Dyhring, T.; Ahring, P. K.; Meyer, M. D. Synthesis and Structure-Activity Relationship Studies of 3,6-Diazabicyclo[3.2.0]heptanes as Novel $\alpha 4\beta 2$ Nicotinic Acetylcholine Receptor Selective Agonists. *J. Med. Chem.* **2007**, *50*, 5493–5508.
- (51) Palmer, A. M.; Grobbl, B.; Jecke, C.; Brehm, C.; Zimmerman, P. J.; Burr, W.; Feth, M. P.; Simon, W.-A.; Kromer, W. Synthesis and Evaluation of 7H-8,9-Dihydropyrano[2,3-c]imidazo[1,2-a]pyridines as Potassium-Competitive Acid Blockers. *J. Med. Chem.* **2007**, *50*, 6240–6264.
- (52) Aurelio, L.; Valant, C.; Flynn, B. L.; Sexton, P. M.; Christopoulos, A.; Scammells, P. J. Allosteric Modulators of the Adenosine A1 Receptor: Synthesis and Pharmacological Evaluation of 4-Substituted 2-Amino-3-benzoylthiophenes. *J. Med. Chem.* **2009**, *52*, 4543–4547.
- (53) Bunnelle, W. H.; et al. Structure-Activity Studies and Analgesic Efficacy of N-(3-Pyridinyl)-Bridged Bicyclic Diamines, Exceptionally Potent Agonists at Nicotinic Acetylcholine Receptors. *J. Med. Chem.* **2007**, *50*, 3627–3644.
- (54) Higuchi, R. I.; Arienti, K. L.; López, F. J.; Mani, N. S.; Mais, D. E.; Caferro, T. R.; Long, Y. O.; Jones, T. K.; Edwards, J. P.; Zhi, L.; Schrader, W. T.; Negro-Vilar, A.; Marschke, K. B. Novel Series of Potent, Nonsteroidal, Selective Androgen Receptor Modulators Based on 7H-[1,4]Oxazino[3,2-g]-quinolin-7-ones. *J. Med. Chem.* **2007**, *50*, 2486–2496.
- (55) Biswas, K.; et al. Potent Nonpeptide Antagonists of the Bradykinin B1 Receptor: Structure-Activity Relationship Studies with Novel Diaminochroman Carboxamides. *J. Med. Chem.* **2007**, *50*, 2200–2212.
- (56) Ménard, D.; et al. Novel Potent BRAF Inhibitors: Toward 1 nM Compounds through Optimization of the Central Phenyl Ring. *J. Med. Chem.* **2009**, *52*, 3881–3891.
- (57) Vachal, P.; adn, T. M.; Fong, J. M. F.; Huang, C. C. R.-R. 1-Sulfonyl-4-acylpiperazines and Selective Cannabinoid-1 Receptor (CB1R) Inverse Agonists for the Treatment of Obesity. *J. Med. Chem.* **2009**, *52*, 2550–2558.
- (58) Tsou, H.-R.; Otteng, M.; Tran, T.; M. B., F., Jr.; Reigh, M.; Birnberg, G.; Kutterer, K.; Ayral-Kaloustian, S.; Rvi, M.; Nilakantan, R.; Grillo, M.; McGinnis, J. P.; Rabindran, S. K. 4-(Phenylaminomethylene)-isoquinoline-1,3(2H,4H)-diones as Potent and Selective Inhibitors of the Cyclin-Dependent Kinase 4 (CDK4). *J. Med. Chem.* **2008**, *51*, 3507–3525.
- (59) Hyatt, J. L.; Moak, T.; Hatfield, M. J.; Tsurkan, L.; Edwards, C. C.; Wierdl, M.; Danks, M. K.; Wadkins, R. M.; Potter, P. M. Selective Inhibition of Carboxylesterases by Istatins, Indole-2,3-diones. *J. Med. Chem.* **2007**, *50*, 1876–1885.
- (60) Sarver, R.; Bills, E.; Bolton, G.; Bratton, L. D.; Caspers, N. L.; Dunbar, J. B.; Harris, M. S.; Hutchings, R. H.; Kennedy, R. M.; Larsen, S. D.; Pavlovsky, A.; Pfefferkorn, J. A.; Bainbridge, G. Thermodynamic and Structure Guided Design of Statin Based Inhibitors of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase. *J. Med. Chem.* **2008**, *51*, 3804–3813.
- (61) Gilligan, P. J.; et al. 8-(4-Methoxyphenyl)pyrazolo[1,5-a]-1,3,5-triazines: Selective and Centrally Active Corticotropin-Releasing Factor Receptor-1 (CRF1) Antagonists. *J. Med. Chem.* **2009**, *52*, 3073–3083.
- (62) Perez-Medrano, A.; Donnelly-Roberts, D. L.; Honore, P.; Hsieh, G. C.; Namovic, M. T.; Peddi, S.; Shuai, Q.; Wang, Y.; Faltynek, C. R.; Jarvis, M. F.; Carroll, W. A. Discovery and Biological Evaluation of Novel Cyanoguanidine P2X₇ Antagonists with Analgesic Activity in Rat Model of Neuropathic Pain. *J. Med. Chem.* **2009**, *52*, 3366–3376.
- (63) Song, Y.; Lin, F.-Y.; Yin, F.; Hensler, M.; Poveda, C. A. R.; Mukkamala, D.; Cao, R.; Wang, H.; Morita, C. T.; Pacanowska, D. G.; Nizet, V.; Oldfield, E. Phosphonosulfonates Are Potent, Selective Inhibitors of Dehydrosqualine Synthase and Staphyloxanthin Biosynthesis in *Staphylococcus aureus*. *J. Med. Chem.* **2009**, *52*, 976–988.
- (64) Wydysh, E. A.; Medghalchi, S. M.; Vadlamundi, A.; Townsend, C. A. Design and Synthesis of Small Molecule Glycerol 3-Phosphate Acyltransferase Inhibitors. *J. Med. Chem.* **2009**, *52*, 3317–3327.
- (65) Chen, K. X.; et al. Second-Generation Highly Potent and Selective Inhibitors of the Hepatitis C Virus NS3 Serine Protease. *J. Med. Chem.* **2009**, *52*, 1370–1379.
- (66) Manfroni, G.; Paeshuyse, J.; Massari, S.; Zanolli, S.; Gatto, B.; Maga, G.; Tabarrini, O.; Cecchetti, V.; Fravolini, A.; Neyts, J. Inhibition of Subgenomic Hepatitis C Virus RNA Replication by Acridone Derivatives: Identification of an NS3 Helicase Inhibitor. *J. Med. Chem.* **2009**, *52*, 3354–3365.
- (67) Nagase, T.; Mizutani, T.; Sekino, E.; Ishikawa, S.; Ito, S.; Mitobe, Y.; Miyamoto, Y.; Yoshimoto, R.; Tanaka, T.; Ishihara, A.; Takenaga, N.; Tokita, S.; Sato, N. Synthesis and Evaluation of Structurally Constrained Quinazolinone Derivatives as Potent and Selective Histamine H3 Receptor Inverse Agonists. *J. Med. Chem.* **2008**, *51*, 6889–6901.
- (68) Ji, L.; Chen, F.-E.; Clercq, E. D.; Balzarini, J.; Pannecoque, C. Synthesis and Anti-HIV-1 Activity Evaluation of 5-Alkyl-2-alkylthio-6-(arylcarbonyl or α -cyanoaryl methyl)-3,4-dihydropyrimidin-4(3H)-ones as Novel Non-nucleoside HIV-1 Reverse Transcriptase Inhibitors. *J. Med. Chem.* **2007**, *50*, 1778–1786.
- (69) Bruno, O.; Brullo, C.; Bondavalli, F.; Schenone, S.; Ranise, A.; Arduino, N.; Bertolotto, M. B.; Montecucco, F.; Ottonello, L.; Dallegri, F.; Tognolini, M.; Ballabeni, V.; Bertoni, S.; Barocelli, E. Synthesis and Biological Evaluation of N-Pyrazolyl-N'-alkyl/benzyl/phenylureas: a

New Class of Potent Inhibitors of Interleukin 8-Induced Neutrophil Chemotaxis. *J. Med. Chem.* **2007**, *50*, 3618–3626.

(70) Procopiou, P. A.; et al. 4-Acyl-1-(4-aminoalkoxyphenyl)-2-ketopiperazines as a Novel Class of Non-Brain-Penetrant Histamine H3 Receptor Antagonists. *J. Med. Chem.* **2007**, *50*, 6706–6717.

(71) Ebdrup, S.; Refsgaard, H. H. F.; Fledelius, C.; Jacobsen, P. Synthesis and Structure-Activity Relationship for a Novel Class of Potent and Selective Carbamate-Based Inhibitors of Hormone Selective Lipase with Acute In Vivo Antilipolytic Effects. *J. Med. Chem.* **2007**, *50*, 5449–5456.

(72) Chimenti, F.; Secci, D.; Bolasco, A.; Chimenti, P.; Bizzarri, B.; Granese, A.; Carradori, S.; nez, M. Y.; Orallo, F.; Ortuso, F.; Alcaro, S. Synthesis, Molecular Modeling, and Selective Inhibitory Activity against Human Monoamine Oxidases of 3-Carboxamido-7-Substituted Coumarins. *J. Med. Chem.* **2009**, *52*, 1935–1942.

(73) Spicer, J. A.; et al. 4-Anilino-5-carboxamido-2-pyridone Derivatives as Noncompetitive Inhibitors of Mitogen-Activated Protein Kinase Kinase. *J. Med. Chem.* **2007**, *50*, 5090–5102.

(74) Anderson, D. R.; Meyers, M. J.; Vernier, W. F.; Mahoney, M. W.; Kurumbail, R. G.; Caspers, N.; Poda, G. I.; Schindler, J. F.; Reinz, D. B.; Mourey, R. J. Pyrrolopyridine Inhibitors of Mitogen-Activated Protein Kinase-Activated Protein Kinase 2 (MK-2). *J. Med. Chem.* **2007**, *50*, 2647–2654.

(75) Sriram, D.; Senthilkumar, P.; Dinakaran, M.; Yogeeswari, P.; China, A.; Nagaraja, V. Antimycobacterial Activities of Novel 1-(Cyclopropyl/*tert*-butyl/4-fluorophenyl)-1,4-dihydro-6-nitro-4-oxo-7-(substituted secondary amino)-1,8-naphthyridine-3-carboxylic Acid. *J. Med. Chem.* **2007**, *50*, 6232–6239.

(76) Qiu, X.-L.; Li, G.; Wu, G.; Zhu, J.; Zhou, L.; Chen, P.-L.; Chamberlin, A. R.; Lee, W.-H. Synthesis and Biological Evaluation of a Series of Novel Inhibitor of Nek2/Hec1 Analogues. *J. Med. Chem.* **2009**, *52*, 1757–1767.

(77) Fish, P. V.; Allan, G. A.; Bailey, S.; Blagg, J.; Butt, R.; Collis, M. G.; Greiling, D.; James, K.; Kendall, J.; McElroy, A.; McCleverty, D.; Reed, C.; Webster, R.; Whitlock, G. A. Potent and Selective Nonpeptidic Inhibitors of Procollagen C-Proteinase. *J. Med. Chem.* **2007**, *50*, 3442–3456.

(78) Choi, S.; Vilaira, G.; Marcinkiewicz, C.; Winkler, J. D.; Bennett, J. S.; DeGrado, W. F. Small Molecule Inhibitors of Integrin $\alpha 2\beta 1$. *J. Med. Chem.* **2007**, *50*, 5457–5462.

(79) Bakunova, S. M.; Bakunov, S. A.; Wenzler, T.; Barszcz, T.; Werbovetz, K. A.; Brun, R.; Hall, J. E.; Tidwell, R. R. Synthesis and in Vivo Antiprotozoal Activity of Bisbenzofuran Cations. *J. Med. Chem.* **2007**, *50*, 5807–5823.

(80) Liu, G.; et al. Discovery of Potent, Selective, Orally Bioavailable Stearyl-CoA Desaturase 1 Inhibitors. *J. Med. Chem.* **2007**, *50*, 3086–3100.

(81) Mandal, P. K.; Ren, Z.; Chen, X.; Xiong, C.; McMurray, J. S. Structure-Affinity Relationships of Glutamine Mimics Incorporated into Phosphopeptides Targeted to the SH2 Domain of Signal Transducer and Activator of Transcription 3. *J. Med. Chem.* **2009**, *52*, 6126–6141.

(82) Yea, C. M.; et al. New Benzylureas as a Novel Series of Potent, Nonpeptidic Vasopressin V2 Receptor Agonists. *J. Med. Chem.* **2008**, *51*, 8124–8134.

(83) Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, S. K.; Hansch, C. Comparative Quantitative Structure-Activity Relationship Studies on Anti-HIV Drugs. *Chem. Rev.* **1999**, *99*, 3525–3602.