Journal of Computer-Aided Molecular Design **17:** 83–92, 2003.

In silico ADME/Tox: why models fail

Terry R. Stouch[1,*], James R. Kenyon[3], Stephen R. Johnson[1], Xue-Qing Chen[2], Arthur Doweyko[1] & Yi Li[1]

[1]*Department of Macromolecular Structure;* [2]*Preclinical Optimization;* [3]*Non-clinical Biostatistics; Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, NJ, USA*

MS received 23 December 2002; accepted 4 January 2003

Key words: ADME, QSAR, Validation, Drug Design, Solubility, HERG, Permeation, 2D6, *In silico*

Summary

By way of example, we discuss the apparent 'failure' of *in silico* ADME/Tox models and attempt to understand the causes. Often, the interpretation of the success of models lies in their use and the expectations of the user. Other times, models are, in fact, of little value. Disappointing results can be linked to the key aspects of the model and modeling procedure, many of these related to the original data and its interpretation. We make recommendations to providers of models regarding the development, description, and use of models as well as the data and information that are important to understanding a model's quality and scope of use.

Introduction

The tremendous – and apparently increasing – cost of pharmaceutical development is common knowledge as is the fact that for every compound that attains the marketplace, at least nine other drug Discovery programs have failed for some reason. The later these failures are identified the higher the cost. Often the liabilities affecting our compounds are not identified until a compound reaches the clinic, the most expensive phase of Pharma R&D.

There is reason to believe that a liability, especially a bad one, will follow a compound throughout its lifespan. Consequently, the earlier a problem is identified, the less time and expense will be invested in an ultimate failure. The reaction to a problem can range from rejection of a series with liabilities, to attempts to 'engineer' them out of an efficacious chemotype, to flagging potential problems for further scrutiny and early detailed in vivo examination.

The liabilities in question include non-optimum values of absorption, distribution, metabolism, and excretion, as well as toxicity and are usually referred to as ADME/Tox or ADMET. Implicit in these properties, but often not explicitly named is physicochemical properties such as solubility, ionization state, crystal forms, surfactant propensity, etc. When physicochemical properties are explicitly included, the acronym becomes PADMET.

Over the last few years, a number of *in vitro* and high-throughput methods have been developed to provide experimental evaluation of these properties as early and rapidly as possible and utilizing as little compound as possible. However, as mentioned above, the earliest possible identification of a liability is desired, including prior even to synthesis. Furthermore, for some assays, the time, throughput, or compound requirements make it impossible to make all of the desired measurements on all desired compounds.

Consequently, there is substantial interest in 'structure-based' theoretical estimates of these properties, often referred to as *in silico* estimates in reference to their computational, as compared to *in vitro* or *in vivo*, nature. The hope is that the *in silico* estimates can be applied to hypothetical compounds, and so guide synthetic efforts. Additionally, there is reason to believe that in some cases the computation can be done quickly and so help to prioritize the large numbers of compounds accessible to combinatorial chemistry. Further, even low-throughput computa-

*Corresponding author. E-mail: Terry.Stouch@BMS.com

tional approaches could be of use in prioritization of compounds for assays that are of limited capacity or that require large quantities of compound.

Given certain circumstances, outlined below, obtaining the *in silico* models from vendors has substantial appeal. However, wherever developed, it must be verified that the models will be useful and usable for the needs of their users.

The requirements of such a model will vary somewhat with user, property, and stage of the effort. However, in our evaluation one constant for all models is that the results must be compatible with in-house experimental estimates of those properties. This point might sustain brief discussion at some stages of an effort, but is irrefutable when an *in silico* method is used to prioritize compounds for experimental assay. Further, it makes sense that evaluation of the utility of a model will make use of in-house data.

Unfortunately, the development of models is not simple. There is a widespread realization that despite the best efforts of their developers, some models that appear to be credible do not have the desired value. This is disappointing to all. However, on investigation and reflection, a number of understandable reasons for this can be identified.

To demonstrate this, we discuss our in-house validation of four models developed by an external vendor. We have reason to believe that the models were sound and of good quality, being done by credible, experienced scientists known to us who had sufficient resources to do a good job and, in most cases, consistent in-house data. According to the developers, the models behaved reasonably during their development and were internally consistent. To their credit, the Vendor did not overstate the value of the models and the modelers were open to discussion and modification of the models. Consequently, this might be considered to be a best case scenario. Despite this we found some of the models lacking for our needs. We probe the reasons for this and give some likely possibilities.

Case histories of difficulties of model validation

Case 1: Data Quality: HERG model

Prolonged QTc has attracted a great deal of attention recently as a precursor to Torsade de Pointes. The inhibition of the hERG potassium ion channel has been implicated as the primary cause of prolonged QTc. It is therefore no surprise that many have developed in silico models to estimate the inhibition of the hERG channel by potential therapeutics.

The model delivered by the Vendor was trained using 76 compounds selected from the literature. The model is a multiple linear regression model of 8 two-point pharmacophoric features.
The data was collected in several different laboratories using a wide variety of assay conditions. For example, the assays used a wide array of cell-types (e.g. HEK, CHO, XO, myocytes) for the expression of the hERG channel. It is also likely that different activation potentials were utilized in different laboratories. It was known that some portion of the training data was binding data rather than inhibition.

A collection of 116 compounds with measured IC_{50}s in the BMS internal patch-clamp electrophysiology experiment were used to validate the vendor model. These measurements were judged to be the highest quality hERG inhibition data available to us for validation. The predictions from the vendor model had a correlation coefficient (R^2) of 0.01 and a root mean square error of 1.3 log units. Clearly, this result demonstrates that the model has no significant relationship to the hERG inhibition data measured in our laboratory ($p \sim 0.23$) (Figure 1).

Our own attempts to utilize the literature data to model hERG inhibition yielded similar results. This is likely due to the widely varying experimental conditions used to collect this data. For example, it is well known that even slight changes in the activation potential can cause widely varying assay results. In addition, the use of non-mammalian cell lines, while possibly somewhat easier to work with experimentally, introduces another source of data variation that is not linked directly to the structures of the potential drugs. As such, it is unrealistic to expect the QSAR modeling to perform well in these conditions.

Case 2: Compound Space: Caco2 passive permeation

In the second case, we evaluated a 21-parameter linear regression model of passive permeation. The model was based on 800 marketed drugs and was a linear regression model based on proprietary pharmacophores. The Vendor using a consistent in-house assay had generated all of the training-set data. At that time, the in-house BMS assay for passive permeation used a different cell-line than did the Vendor. Realizing the possible inconsistency in the two experiments, values were compared between the assays for 40 standards. The agreement between the assays was substantial,

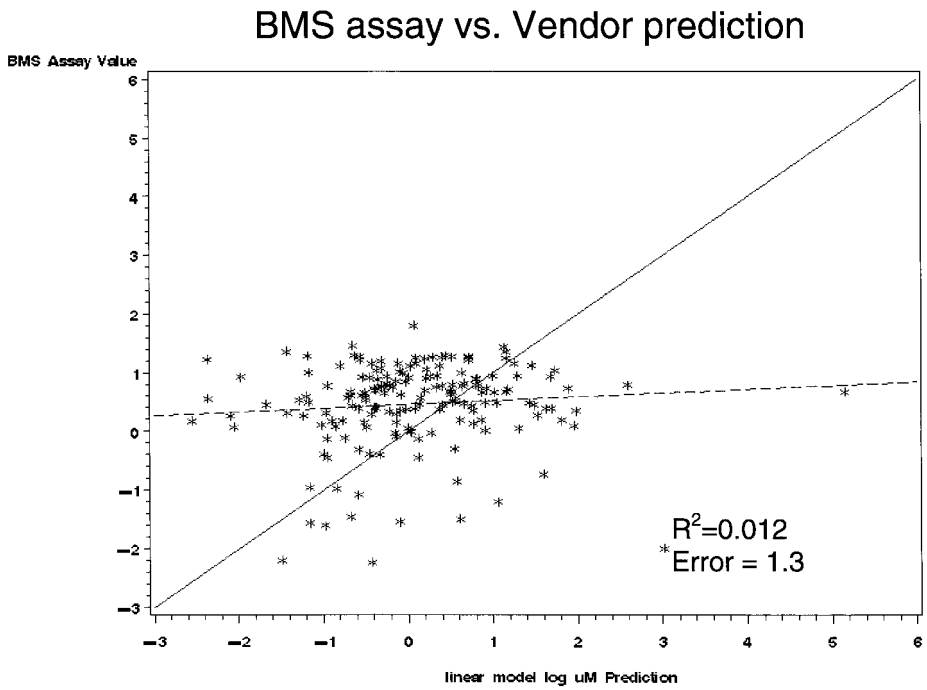

Figure 1. hERG inhibition: BMS assay vs. Vendor model prediction. Solid line denotes ideal, dotted line actual correlation.

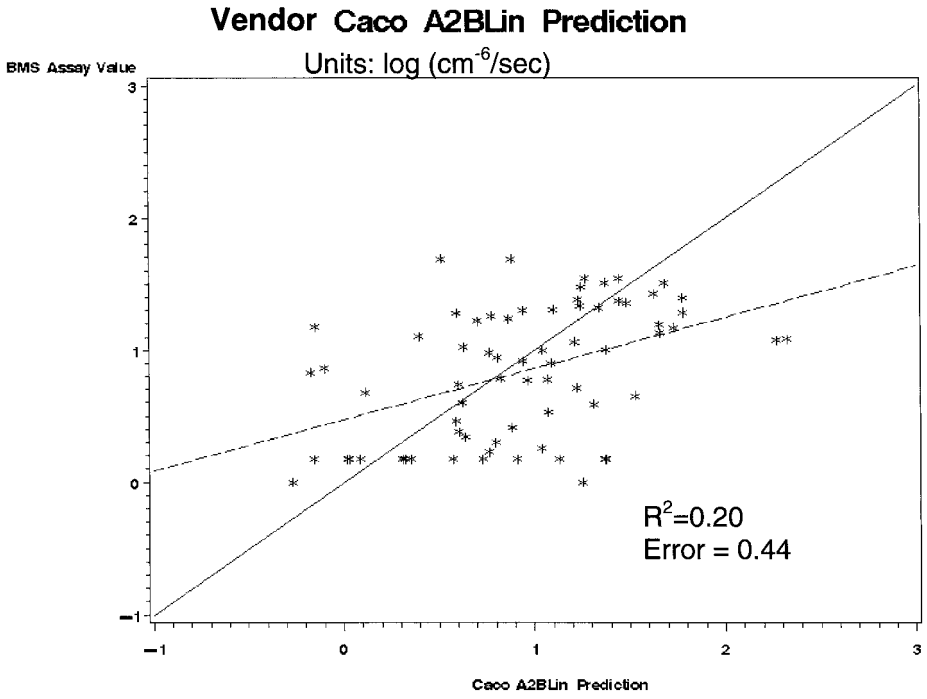

Figure 2. Passive permeation: BMS CACO-2 A to B assay vs. Vendor model prediction. Solid line denotes ideal, dotted line actual correlation.

almost quantitative. Consequently, we had reason to believe a model resulting from the vendor's data would be applicable to of BMS.

We applied the model to 76 BMS compounds containing a range of structural types drawn from a number of Discovery programs and assayed over a number of years. The data was selected to be of the highest available quality with established high recoveries.

In brief the results of the model evaluation are shown in Figure 2. The correlation coefficient of 0.2 is obviously poor. Discrepancies in the experimental values between the Vendor and BMS for the standard compounds could be rationalized based on differential expression of transporter proteins. However, this could not account for the discrepancy between the BMS assay results and the Vendor model prediction show in Figure 2.

In a retrospective analysis, the vendor made available the structures of the compounds used in generation of the model. Consequently we were able to do a direct similarity comparison between the basis of the model and the test compounds. Few of the parities 2-dimensional similarities (Daylight keys) were greater than 0.3 on a scale of 0–1. On this scale, 0 implies no similarity whatsoever (a rare event) and 1 implies identity. Values of 0.9 identify close homologues and those less than 0.5 show little apparent structural similarity.

Consequently, it might be asked if it was appropriate for us to apply the model to this set of compounds. Note that the model was based largely on molecular pharmacophores which would identify spatial arrangements of significant chemical moieties (acid, base, hydrophobe, hydrogen bond donor/acceptor, etc.). Consequently, the model would be dependent on the molecular chemotypes and 3-dimensional structures employed by both the modeling and test procedures. (In our case, the 3-dimensional structures were generated by separate procedures using the 'standard' in-house protocol of the default CONCORD generated structure [1]). Models based on this type of descriptor, encompassing explicit structural features and 3-dimensional information, would be more prone to the importance of molecular similarity than one based on 'general' molecular physicochemical properties such as size, shape, logP, etc.

Perhaps the most surprising result of this analysis is not the poor prediction, but that a reasonably large training set of 800 drugs was so dissimilar to compounds directly under investigation as putative drugs. However, Lipinski has been instrumental in the exposition of the recent trends of pharmaceutical structure and properties [2].

Based on our findings, the final evaluation of this model was not that it was a 'failure' or somehow inadequate, but simply that it was not sufficient for current BMS needs in that it did not cover the necessary range of chemical space. Had we the luxury of continuing the relationship with the Vendor, it would have been instructive to at least re-train the model with BMS data. It is likely, however, the choice of descriptors might entail the addition of new pharmacophores to describe new structural features. If this was so, in our opinion it would be a shortcoming in the model, since it would not have been truly general in the identification of important calculable *properties.*

As will be noted, below, a drawback toward continued development of this particular model is that the Vendor used proprietary pharmacophoric descriptors and software for this model. These are unavailable to us outside of a particular arrangement with this Vendor. Consequently, the long-term use of a resulting reparameterized model would not be guaranteed.

Case 3: Model Endpoint: Cyp 2D6 inhibition

In a third case, a model of Cytochrome P450 2D6 inhibition was evaluated. The 15 parameter (primarily pharmacophores) linear regression model was based on almost 250 drugs and drug-like compounds recently assayed by the Vendor in a consistent in-house assay. Based on the known accuracy of the assay, the modeling had modest goals, classification of compounds as inhibitors or non-inhibitors, using a training set cutoff of 10 μM.

This model was applied to almost 600 BMS compounds assayed internally over five years and from approximately 100 different Discovery programs. Figure 3 shows one interpretation of the data. Here, each bar represents a particular range of inhibitory activity as determined by a BMS-internal assay. The coloration depicts the number of those compounds that the model predicted to be either active (lower bar) or inactive (upper bar). For example the first bar represents the structures that were determined experimentally be non-inhibitors at 50 μM or above. For this range of activity, the model was correct approximately 90% of the time. Indeed, for the remainder of the classes, the predicted results appear to follow reasonable trends.

However, the primary users of a model of this sort want to evaluate a compound more precisely and the BMS-determined alert value is 10 μM. Figure 4

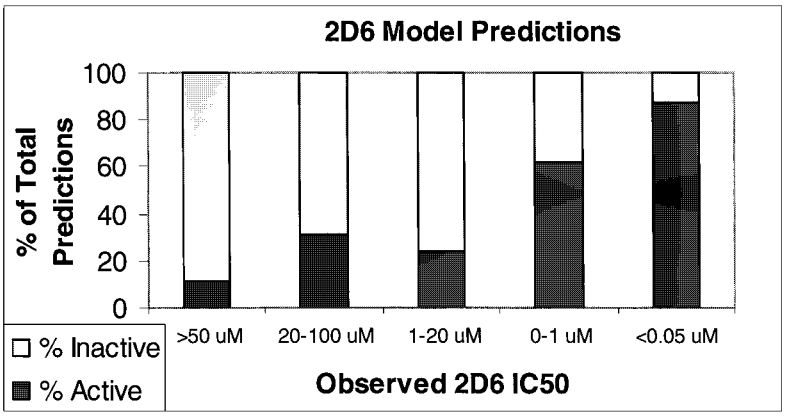

Figure 3. Cytochrome P450 2D6 inhibition: The dark, lower, bars show the percentage of the compounds in that activity range predicted by the model to be either active or inactive.

shows a breakdown of the classifications based on a predicted activity of this value. At this level of resolution, the model's success is less favorable. 'Active' compounds with known inhibition of 10 μM or less were predicted correctly on 60% of the time; 40% of the predictions were falsely predicted as being non-inhibitors. A primary use of such a model would be to identify compounds likely to have a liability and hence subjected to further examination, such as experimental assay. Consequently, the identification of 40% of the known inhibitors to be benign is troubling. Less critical but still problematical is the identification of 30% of the inactive compounds to be active, which would raise concerns falsely and could result in an ineffective use of resources as well as diminishing confidence in the model. A more detailed analysis failed to find substantial benefit beyond the classification of very benign and very malignant compounds.

Despite these observations, the model can not be faulted for use at a resolution below that for which it was designed. However, it was not precise enough to be of use for the majority of users at BMS. It might, however, be of value in screening virtual libraries or ranking results of high-throughput screening. One substantial complaint about this model, however, is that it gave a quantitative estimate of the prediction to an inappropriate precision.

Indeed, a common criticism of vendor-supplied models for predictive ADMET is a general disregard for even the simple concept of significant figures. While a seemingly minor detail, the lack of consideration of how many figures are significant implies little forethought of the quality of the data used to derive the model. It is our experience that if you deliver 4 significant figures to our chemists they will ultimately utilize them to rank compounds for consideration. In cases where data is only accurate to 2- or 3-fold, often the case with high-throughput IC_{50} data, it is misleading to provide 4 or 5 'significant' figures simply because the regression method returns them. The lack of consideration of data quality and error becomes even more important when continuous data are 'binned' into discrete categories. Compounds close to the category cutoffs are problematic as the experimental error could quite easily dictate that the compounds equally belong to multiple classes.

Case 4: Assay agreement and data range: solubility measurements

In the fourth and final example, a solubility model was tested. The Vendor developed a 19 parameter (primarily pharmacophore) linear regression model based on over 1000 marketed drugs with solubilities ranging from 1–200 μM using an in-house medium-throughput (MT) assay involving precipitation from a concentrated DMSO solution. The validation set was 533 diverse BMS compounds assayed in a similar MT assay over 3 years with solubilities ranging from 0.1 μM to 600 μM. The model provided results as classification as insoluble, moderately soluble, or very soluble as shown in Figure 5. Although a little difficult to interpret, the overall conclusion is that the predicted results were little better than random.

On closer examination of the data and experimental protocols it was determined that assay conditions differed slightly between the Vendor and BMS in terms of pH, shake times, and equilibration. Although those running the assays felt that these were

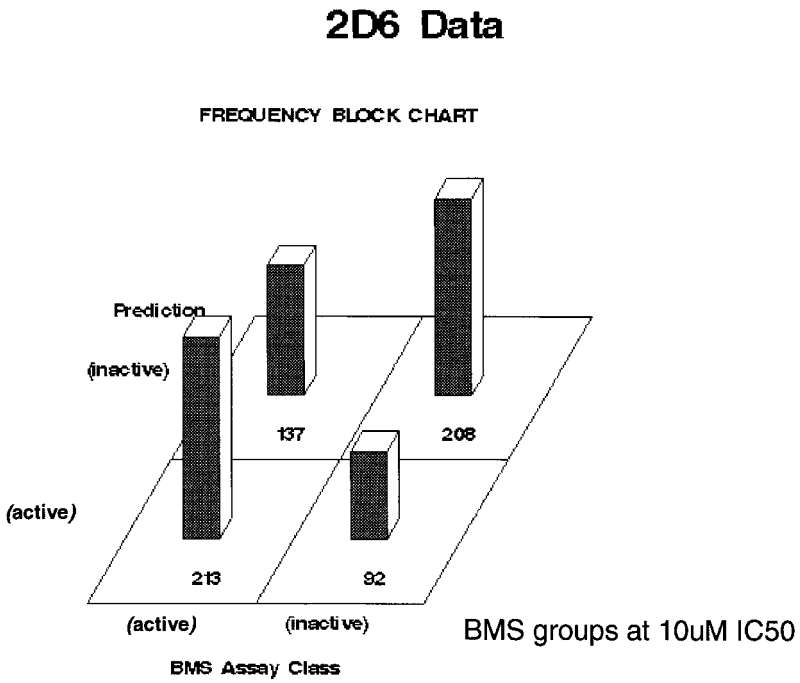

Figure 4. Cytochrome P450 2D6 inhibition: A different representation of the predictions this time classified by the BMS alert cutoff of 10 μM.

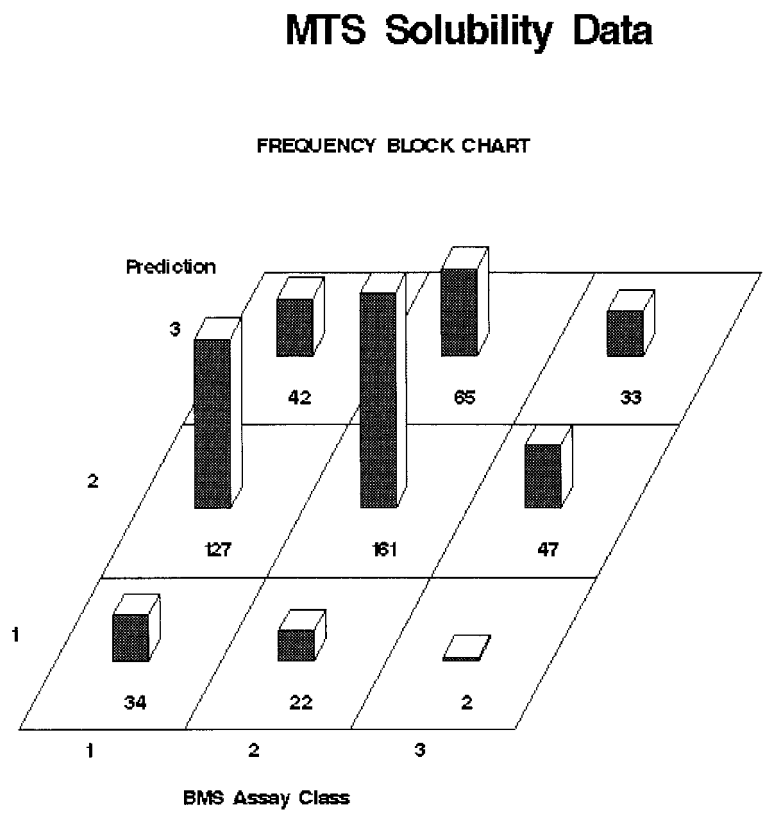

Figure 5. Solubility prediction of BMS medium throughput assay data subdivided into low, medium, and high solubilities.

minor differences considering the overall error in the assays, this was not established. Indeed, it is this overall error that is of particular concern. This type of MT solubility assay has proven to be problematical and of modest agreement with more labor-intensive equilibrium solubility assays.

An additional potential problem with the model was the small range of solubility used in development of the model. A 1μM lower limit is, sadly, a solubility value to which only a few welcome Discovery compounds can aspire, much less exceed. The 200 μM ceiling was the upper limit of their assay. Consequently, the limited data range could have compromised the model parameterization. None-the-less, the needs of pharmaceutical research go well beyond the range probed by the training set or the test set employed here.

This raises an important concern: what is the true endpoint of the assay? Solubility can mean many things and one needs to be cognizant of the exact endpoint of the data from which the model was developed. Although not applicable in this case, of especial concern is that some modeling efforts, as noted above for the HERG model, derive data from many different sources with the concomitant probability of differences not only between assays, but in definition of the endpoints.

Advice to vendors

We have examined four cases of model 'failure': inadequate data, insufficient range of compound space, inappropriate endpoint, and different definitions of the same endpoint. We were fortunate to have a relationship with the vendor (in our experience, unique) that allowed us to uncover these issues. Had it not been for that, we would have concluded that the models were of little value and might have questioned the competence of the modeling effort, despite our favorable personal judgements.

From evaluations of models such as these, we have learned valuable lessons in the interpretation and use of models. These lessons influence our own model development and we would like to convey them to those who would provide models to us. This information is vital to the proper evaluation of a model and can make a distinct difference in the ultimate perception of a model's value.

Tell us about the data

The training and test set data is the most important part of a model. The data source(s) should be clearly disclosed along with references, if appropriate. Details of the assays should be disclosed as well as the error estimates of the values and the true nature of reported endpoints. The degree of confidence in the data point should be reported, if available.

The quantity and diversity of the data should be represented: how many compounds were used, the structural range and similarity of the compounds, the types of compounds that were not represented. The diversity is important both within chemical spaces as well as within the space of the model. A very important consideration is how the training set compares with the targeted use of the model. Regarding this point, we recommend use of Pearlman's Diverse Solutions, BCUTS, estimate of data diversity for comparison of a training set with known drug space or our in-house data space [3, 4]

Of particular importance are the compounds in the training set. Without these, we can not evaluate a comparison of the chemical scope of the model with the compounds to which it will be applied. Although this comparison can in part be determined via comparisons of a DVS space (data 'spaces' can be shared without revealing the structures of the underlying compounds) little can compare with a compound-by-compound comparison. In particular, as noted above, it is important to understand the similarity within the space defined by the descriptors comprising the model in addition to the more global metrics, such as DVS. There is substantial value to knowing the structure, identity, experimental value, and predicted value of that compound in the training set most similar to the predicted compound. *This knowledge will impact the interpretation of the value of the model*, as noted above in the validation of the passive permeation model. Additional value comes from mining the training set for rules and trends that might be used to guide synthesis. For example, how do structural variations of a chemotype affect the liabilities in question?

The assembly of a quality training set takes time and effort and it is understandable that release of this publicly might compromise a Vendor's intellectual property and competitive advantage. However, IP protection by legal agreement is precedented within Pharma.

Tell us about the model

Details of the model and the modeling procedure should be disclosed. For example, the nature of the model: linear regression, PLS, clustering, discrimination, other. The treatment of the experimental values is important. Have they been used as quantitative variables? Have they been segregated into classes? If so, how and why? The consideration of experimental error during the modeling and in the interpretation of the prediction is important.

The success of the modeling procedure should be expounded. The statistical details of the modeling procedure should be discussed as well as detailed of the predictive capabilities of the model. For example, not only the rate of correct prediction but also that of erroneous prediction. Comparison of this with random expectation should always be made. Was this for particular structural types, within particular ranges of activity, or most evident for compounds of particular properties? In other words, the true range and applicability of the model should be investigated in every way possible.

The descriptors that were examined as well as those that participate in the final model are important information. If appropriate, the contribution and weighting of each should be made clear. Do these make sense? How does this relate to the property in question? If novel descriptors are used, explain why. Do they in fact provide additional information over better-known, more accessible descriptors? Are the differences significant? One might question the commercial value of the dependence on proprietary software or descriptors over more accessible models, even if the predictive value suffer somewhat.

Does the model make physical or biological sense? Although this is not an absolute requirement for a truly validated model with demonstrated predictive capability, it does provide reassurance when a model has anything but a stellar record. Further, a prediction of a value or activity is only one concern. Almost as important is an understanding of the basis of that value.

The predicted values/residuals/classifications of the training set compounds is an important diagnostic. Additionally, if training set compounds did not fit the model, or were left out of the analysis the justification should be explained.

How does the model compare with other models with the same endpoint? Is the difference really significant? The literature is replete with examples of statistically insignificant improvements in models of identical training sets.

In all cases, unless a modeling effort is quite certain of the generality of the model, it must be tailorable to allow the inclusion of new data. This must be doable with consideration of the potential proprietary nature of the new data. Further, the procedure for doing this should allow for a complete statistical evaluation of the resulting refit model.

Finally, the developer of a model should have and convey a clear sense of who is the intended primary user of a model. Can a chemist intent on improving solubility use it by a factor of two? Or is it best used in the hands of HTS prioritization, where estimates within a few orders of magnitude might be acceptable?

Help us to understand the model and the property

The goals of ADME/Tox modeling are not only to identify a liability, but also to understand its cause and how liabilities might be engineered out of a promising chemotype. In addition to an estimate of the value of the liabilities, often additional information is available that could help with this process.

For example, in some cases recommendations might be made regarding approaches to vary the value of the endpoint. Problematic structural features might be identified. Examples of the effects of variations of similar compound could provide guidance.

If the model and the descriptors allow for an interpretation of the physical basis for the property, it could be explicitly highlighted. Would varying hydrophobicity help? Increasing electron density in a key area? Decreasing bulk at a particular point in space?

Whenever possible, it is reassuring and useful if QSARs can be related to 3-d compound/protein interactions. For example, QSAR and pharmacophoric models of the Cytochrome P450s have been reported. In addition, a mapping of this information onto structures or homology model could provide an intuitive guide for synthesis.

ADMET is a tough field and every bit of information often needs to be brought to bear. No doubt the careful ADMET modeler will have canvassed the literature. It would often be useful if some of these additional sources of information could be made available to the user in addition to simple predictions. These might include literature references and databases as well as key contract firms and leading academic labs.

The training set, alone, could provide useful ‘additional’ information. For example, it might be mined to

uncover how particular variations in structure affects a particular liability. For a given compound, what is the most similar compound with a desirable value of the property? An undesirable value? How large a change in structure might elicit a desired effect of property? As noted, above, a direct compound-by-compound comparison of the test compound with the most similar compound in the training set would provide valuable information on the confidence of the prediction.

Discussion

It is becoming increasingly evident that *in silico* ADME/Tox modeling is a difficult prospect. On close examination and use, many models either fail or appear to fail. Perhaps this should not be surprising. How much do we expect for a QSAR within a Drug Discovery optimization effort employing a single target, a quantitative assay, a single laboratory, and one or a few related chemotypes? Yet, expectations for ADME/Tox modeling are as great or greater.

Its seems that often a QSAR/QSPR of ADME/Tox properties is expected to be able to handle most any type of compound with in one model or a limited set of models. This might not be an unreasonable goal for properties dependent on general molecular physical properties, such as passive diffusion across lipid bilayer membranes. However, for more complex biological endpoints, such as carcinogenesis, which is the result of multiple mechanisms, the hope exceeds the realities.

The data that has been used in ADME/Tox modeling can represent an important problem that perhaps can be remedied. Often this data is hard-won. Consequently, modeling efforts have had to resort to using available data from where ever it is available. Usually that comes from multiple laboratories with the inherent incompatibilities that accompany it. Data such as this should be carefully critiqued. As in the case of the hERG inhibition model, above, it might not even be worth using. This is an important consideration to which we must be present. Availability of data does not necessarily mean that a model can or should be developed.

Recent advances in high-throughput assays now provide the possibility of more consistent data. The goals, endpoints, and accuracy of these assays must be carefully considered, however. For example, is the goal to qualitatively determine the existence of a liability, or to provide an accurate absolute value of the liability? For example, the goal of a passive permeation assay (CACO-2, MDCK, PAMPA) might be to provide a qualitative estimate of permeability – permeable vs. non-permeable. Consequently, there might be no need to determine values with accuracy greater than or less than target values. The resulting data might be unsuitable or even counterproductive to a modeling effort.

As another example, due to its automated nature high-throughput assays will use predetermined concentrations of test compounds. If these concentrations do not sufficiently populate a dose-response curve, a resulting IC50 could be compromised. For these reasons, it is not unusual for high-throughput data to be 'binned' in ranges for interpretation or comparison with other assays. If this binning approach is representative of the confidence in the numerical measurements, they should not be interpreted to greater precision when used for *in silico* modeling.

In the long run, it is worth the effort and expense to obtain data of a range, precision, and accuracy that can support a modeling effort with the desired endpoint. The properties in question are important, costly, and ever present. The costs involved in procuring this data should be considered to be an amply warranted investment.

Lastly, any QSAR/QSPR effort must be aware of the statistical pitfalls which can occur [5–9]. This is particularly important when data becomes less quantitative and expectations for results are not so great. Results should be scrutinized carefully and careful detailed validation is essential [10]

Not all is gloomy, however. As we progressively gain understanding of the problems and pitfalls, the *in silico* models will improve. Successes are already present to some degree, especially for sets of compounds with limited structural variation. As we understand the data requirements for a particular modeling endpoint, more suitable data can be obtained. As we understand the modeling pitfalls, more care can be taken with development and validation.

Conclusion

As we gain experience with the development and use of ADME/Tox models, problems with their development and use are becoming evident. Often, these problems might make a carefully considered model appear to be of low quality and perhaps useless. In part, this can be remedied by a careful exposition of

the data used to develop the models and a detailed comparison of the test compounds to that data. Endpoints, goals, accuracy, precision conditions of the assays from which the data is derived must be considered and disclosed. Clear exposition of the model and its development are essential.

Although the results from many ADME/Tox models has been disappointing, there is every reason to believe that with continued effort successes will be increasingly prevalent. An important consideration for these efforts is the availability of suitable data. The industry should consider the acquisition of this data to be a worthwhile investment.

Acknowledgements

The authors would like to thank the following for their insight, efforts and support: Saeho Chong, Michael Sinz, Srini Venkatesh, Deborah Loughney, Wesley Cosand, Scott Biller, Karnail Atwal.

References

1. Pearlman, R.S., *Concord User's Manual*, Tripos, Inc.: St. Louis, Mo.
2. Lipinski, C.A., J. Pharm. Tox. Meth. 44 (2000) 235–249.
3. Pearlman, R.S. and Smith, K.M. J. Chem. Inf. Comput. Sci. 39 (1999) 28–35.
4. Pearlman, R.S. and Smith, K.M. Perspect. Drug Discov. Design 9-11 (1998) 339–353.
5. Topliss, J.G. and Edwards, R.P. J. Med. Chem. 22 (1979) 1238–1244.
6. Stouch, T.R. and Jurs, P.C. J. Chem. Inf. Comput. Sci. 25 (1985) 45–50.
7. Stouch, T.R. and Jurs, P.C. J. Chem. Inf. Comput. Sci. 25 (1985) 92–98.
8. Stouch, T.R. and Jurs, P.C. Quant. Struct.-Activity Relat. Pharmacol. Chem. Biol. 5 (1986) 57–61.
9. Clark, R.D., Perspect. Drug Discov. Design, (2003) This volume.
10. Tropsha, A., Gramatica, P. and Gombar, V.K. Quant. Struct.-Activity Relat. (2003). In Press.