

MODELING BIOLOGICAL ACTIVITY

Discovering new drugs takes lots of time. Proving to FDA that a new drug works and is reasonably safe takes even more time. By the time a new drug is brought to market, most of the patent has expired and millions of dollars have been spent with no return on investment. One of the reasons pharmaceuticals are so expensive is the current drug development methodologies create a lot of false leads. Similarly, it costs an outrageous amount to do animal testing for toxic compounds. All drugs have side effects and some are potentially toxic.

In order to more clearly understand the process for modeling biological activity and predicting the interaction between molecules and biological systems, it will be helpful if we use a common problem and follow the process through for a real life example. We will assume for this discussion that we are looking for new molecules that will act as blood pressure lowering drugs and are active biologically as beta-blockers. For this example, let's assume we have thirty known molecules that show varying effectiveness as beta-blockers and lower blood pressure (i.e. the existing drugs used for this purpose today).

QSAR: Most common technique used today

The most common technique for modeling chemical/biological interactions is called Quantitative Structure Activity Relationships (QSARs). The first task for the researcher is to identify a set of chemical structural parameters that define the biological activity of known beta-blocker drugs. Often this results in as many as two hundred or more parameters for each molecule evaluated. In our example, where we have thirty known molecules that are classified as beta-blockers, that can typically take up to a month to perform these calculations. The calculated data is then entered into a computer program and a statistical analysis is performed to define a model that predicts whether a new molecule of interest might be an effective beta-blocker.

At this point the researcher is ready to evaluate new molecules. Let's say the researcher picks thirty new molecules for evaluation. Each molecule must go through the same procedure and all the required parameters must be calculated for each new molecule. This will take more time. At this point each new molecule is analyzed against the model, which predicts the biological interactions. Once a few of the molecules are determined to be potential candidates then they must be synthesized, tested and go through arduous testing with live animals for determination of side effects, efficacy and toxicity.

There are several problems with this technique:

1. Biological systems are active in an aqueous environment; our body is 70% water. Chemicals do change shape in aqueous systems, which will change their chemical interactions and activity, an effect that is also directly dependent on the temperature of the system. QSAR does not directly predict what happens in biological systems.
2. The researcher must make a lot of judgment calls when calculating the parameters or choosing which parameters to use for the model, and it is not an exact science. Even though these are educated judgments they are still fallible and puts a significant element of subjectivity and uncertainty to the effectiveness of the analysis and the model developed to predict the effectiveness of new molecular structures.
3. Another problem with this technique is that the researcher predefines a set of the most critical parameters that are viewed as important for the interactions desired. Only new molecules that fit into this system nicely will be evaluated. If a molecule has good potential but differs in a substantial way from the preset structure it can't be evaluated by the model and is therefore ignored. Often new and promising structures are then outliers and can't be evaluated with this approach.
4. Due to the estimates and human judgments applied in this process the resultant model that is developed may not be optimum for evaluating new molecules. Typically these predictions are 30-40% accurate. So over 60% of the molecules that potentially could go to the expensive route of synthesis, efficacy testing and toxicity testing in animals are dead ends. These facts make the entire effort costly, time-consuming and ineffective. From a business standpoint it is a poor tool.

SDAR: an NCTR Discovery

The SDAR Story: In 1998 Drs. Dwight Miller, Jack Jay, and Jon Wilkes discussed the arbitrariness and computational intensity involved in development and use of QSAR. They wondered whether the problems could be avoided and calculations, simplified by building relationships directly between molecular spectra and biological effects. Dr. Miller suggested that ^{13}C Nuclear Magnetic Resonance (NMR) would provide the best information. His reason for recommending NMR was based partly on the fact that these spectra are acquired in liquid solution so at the time of measurement the molecules are already folded into a Boltzmann distribution of conformations, shapes similar like they would look in the biological systems. Dr. Wilkes checked the ^{13}C Nuclear Magnetic Resonance (NMR) spectra of alpha- and beta-estradiol. (These compounds have the same atoms bound together in the same sequence except that on top of one structure a hydroxyl group tips to the front and on the other it leans to the rear. Yet their estrogenicity differs by a factor of ten. Clearly tiny changes in the

hydroxyl group orientation can make a huge difference in the estrogen effect.) Sure enough, the two ^{13}C -NMR spectra were quite different. This gave reason to believe that pattern recognition based on spectra of strong and weak estrogens could learn to recognize the spectral clues that distinguished the strong and weak interactions. Thus was born the Spectral Data-Activity relationship (SDAR) concept. Dr. Rick Beger, an expert in both NMR and computational methods, came to NCTR and with the others began developing SDAR models of biological effects.

The differences between the current method of QSAR and SDAR are striking. First the spectra can be obtained in liquid solution, just like in real biological systems. There are no human judgments entering into the process since the raw data employed is actual NMR spectra of the specific molecule being evaluated. It also is clear that very small changes in molecular shape clearly show up as major changes in the spectra showing the sensitivity of this technique. Also there are no arbitrary constraints on the process or human judgments in calculating activity parameters to define a model. Outliers are welcome.

For our example, of the thirty known beta-blocker molecules, one obtains the NMR spectra and enters the spectra into specialized software that uses an artificial neural network (or some other pattern recognition algorithm) to develop the model to predict biological reactivity. At this point we are ready to look at new molecules. The NMR spectrum for the target molecule is entered into the model and with the use of pattern recognition software we can predict its effectiveness for lowering blood pressure as a beta-blocker; ie. strong versus medium versus weak interactions. With this technique correlation coefficients greater than 90% are common for predicted activity against experimental data.

QSDAR: Enhancement of SDAR

Eventually, SDAR models were developed using a different computer program so they could give quantitative estimates - not just strong/weak classifications, but numerical estimates of efficacy. This technique is called quantitative SDAR (QSDAR). Another major improvement in QSDAR is the use of computer-predicted, rather than instrumentally acquired spectra, for model building purposes. (It takes a long time to acquire an NMR spectrum, but very little to predict what it should be.) Rather than taking up to a month or two, the entire modeling process can be accomplished by a single analyst working at a PC in about two days.

Applications:

These models can be used to identify therapeutic drug leads, to discover new antibiotics, and to anticipate undesirable side effects such as those recently disclosed for Celebrex™ and Vioxx™. The same methods have been used successfully to predict that certain dioxin-like compounds were much more toxic than expected. This should contribute greatly to quantitative risk assessment of contaminated sites as well as for food safety. This approach can also assist medical researchers in designing pharmaceuticals with activity against bio-engineered bacterial agents, designer “bugs” with multiple drug resistance or multiple attack modes.

A real life example: QSDAR methods were used successfully to model Toxic Equivalency Factors (TEFs) of dioxins, furans, and PCBs¹⁴. The models predicted that several congeners assigned zero TEFs by the USEPA and the World Health Organization actually should have significant toxicity. The predictions for two of the congeners have now been independently confirmed by a commercial method called CALUX.

References:

¹ "Methods for Predicting the Biological, Chemical, and Physical Properties of Molecules from Their Spectral Properties" D. Miller, R. Beger, J. Wilkes, J. Lay, Jr., and J. Freeman. US Patent Application No. 09/496,314, filed in USPTO February 1, 2000. Granted with 74 claims, **June 8, 2004**.

²"Methods for Predicting Properties of Molecules" R. Beger and J. Wilkes, Patent Application 10/383,602 Filed in USPTO **March 07, 2003**.

³ R. Beger, J. Freeman, J. Lay Jr., J. Wilkes, and D. Miller, Producing ¹³C NMR, Infrared Absorption and EI Mass Spectrometric Data Monodechlorination Models of Chlorobenzenes, Chlorophenols, and Chloroanilines *J. Chem. Inf. Comput. Sci.* **2000**, 40:6,1449-1455.

⁴Richard D. Beger, James P. Freeman, Jackson O. Lay Jr., Jon G. Wilkes, Dwight W. Miller, "13C NMR and Electron Ionization Mass Spectrometric Data-Activity Relationship Model of Estrogen Receptor Binding" *Toxicology and Applied Pharmacology* **November 2000** 169:1, 17-25.

⁵ Richard D. Beger, James P. Freeman, Jackson O. Lay, Jr., Jon G. Wilkes, and Dwight W. Miller. "Use of ¹³C NMR Spectrometric Data To Produce a Predictive Model of Estrogen Receptor Binding Activity, *J. Chem. Inf. Comp. Sci.* **2001**, 41:1, 219-224.

⁶ R. Beger, J. Freeman, J. Lay Jr., J. Wilkes, and D. Miller, Producing ¹³C NMR, Infrared Absorption and EI Mass Spectrometric Data Monodechlorination Models of Chlorobenzenes, Chlorophenols, and Chloroanilines *J. Chem. Inf. Comput. Sci.* **2000**, 40:6,1449-1455.

⁷ Richard D. Beger, Dan A. Buzatu, Jon G. Wilkes, and Jackson. O Lay, Jr. "¹³C NMR Quantitative Spectrometric Data-Activity Relationship (QSDAR) Models of Steroid Binding to the Aromatase Enzyme" *J. Chem. Inf. Comp. Sci.* , **(2001)** 41:1360-1366.

⁹ Richard D. Beger, Jon G. Wilkes "Developing 13C NMR quantitative spectrometric data-activity relationship (QSDAR) models of steroid binding to the corticosteroid binding globulin" *Journal of Computer-Aided Molecular Design* incorporating '*Perspectives in Drug Discovery and Design*', **July 2001**, 15:7, 659-669

¹⁰ Richard D. Beger, and Jon G. Wilkes "Models of Polychlorinated Dibenzodioxins, Dibenzofurans, and Biphenyls Binding Affinity to the Aryl Hydrocarbon Receptor Developed using ¹³C NMR Data" *J. Chem. Inf. Comp. Sci.* , **(2001)** 41:1322-1329.

¹¹ Richard D. Beger and Jon G. Wilkes "Comparative structural connectivity spectra analysis (CoScCoSA) models of steroids binding to the aromatase enzyme" *J. Mol. Recog.* **2002**, 15:1-9.

¹² Richard D. Beger, Dan A. Buzatu, Jon G. Wilkes, and Jackson O. Lay "Comparative Structural Connectivity Spectra Analysis (CoSCoSA) Models of Steroid Binding to the Corticosteroid Binding Globulin" *J. Chem. Inf. Comput. Sci.* **2002**, *42*:1123-1131.

¹³ Richard D. Beger, Dan A. Buzatu, and Jon G. Wilkes "Combining NMR spectral and structural data to form models of polychlorinated dibenzodioxins, dibenzofurans, and biphenyls binding to the AhR" *Journal of Computer-Aided Molecular Design*, **October, 2002**, *16:10*, 727-740.

¹⁴ Lindsay Shade, Richard D. Beger, and Jon G. Wilkes "The Use of Carbon Thirteen Nuclear Magnetic Resonance Spectra to Predict Dioxin and Furan Binding Affinities to the Aryl Hydrocarbon Receptor" *Environmental Toxicology and Chemistry*, *22:3*, **2003**, 501-509.

15 B. Hass, L. Pence, J. Archer, R. Beger, J. Wilkes, and D. Buzatu "QSDAR prediction of TEF values for 1,2,3,4,7-PCDD and 1,3,7,8-TCDD validated by CALUX-derived binding values." for submission by **February 1, 2005**, to *Environmental Toxicology and Chemistry*.

¹⁶ Richard D. Beger, Kathleen J. Holm, Dan A. Buzatu, and Jon G. Wilkes, "Using Simulated 2D ¹³C NMR Nearest Neighbor Connectivity Spectral Data Patterns to Model a Diverse Set of Estrogens" *Internet Electron. J. Mol. Des.* **2003**, *2*, 435-453, <http://www.biochempress.com>.

¹⁷ Dan A. Buzatu, Richard D. Beger, Jackson O. Lay, Jr., and Jon G. Wilkes, "Predicting Toxic Equivalence Factors from ¹³C Nuclear Magnetic Resonance Spectra for Dioxins, Furans and PCBs Using Multiple Linear Regression and Artificial Neural Networks", *Environmental Toxicology and Chemistry*, **2004**, *23:1*, 24-31

¹⁸ Jon Wilkes, "NCTR's **SDAR** Models of Biological Activity"