Fuzzy Modeling of Skin Permeability Coefficients[1]

Angela K. Pannier,[2] Rhonda M. Brand,[3] and David D. Jones[2,4]

Purpose. The purpose of this work was to determine whether a new modeling methodology using fuzzy logic can predict skin permeability coefficients that are given compound descriptors that have been proven to affect percutaneous penetration.

Methods. Three fuzzy inference models were developed using subtractive clustering to define natural structures within the data and assign subsequent rules. The numeric parameters describing the rules were refined through the use of an Adaptive Neural Fuzzy Inference System implemented in MatLab. Each model was evaluated using the entire data set. Then predicted outputs were compared to the published experimental data.

Results. All databases produced fuzzy inference models that successfully predicted skin permeability coefficients, with correlation coefficients ranging from 0.83 to 0.97. The lowest correlation coefficient resulted from a model using log octanol/water partition coefficient and molecular weight as inputs with two input membership functions evaluated by two fuzzy rules. The correlation coefficient of 0.97 occurred when log octanol/water partition coefficient and hydrogen bond donor activity were used as inputs with three input membership functions evaluated by three fuzzy rules.

Conclusions. Fuzzy rule-based models are a realistic and promising tool that can be used to successfully model and predict skin permeability coefficients as well as or better than previous algorithms with fewer inputs

KEY WORDS: fuzzy logic; skin permeability; percutaneous absorption; clustering; adaptive neural fuzzy inference system.

INTRODUCTION

Skin permeability is an important parameter in the assessment of potential toxicity of environmental agents or the feasibility of a drug for transdermal delivery. Although skin penetration can be determined experimentally, a simple model that can predict this descriptor, based on few inputs, is invaluable to both risk assessment and drug-delivery investigations. A number of algorithms to predict skin permeability coefficients have been developed, using empirical, analytical, and theoretical approaches (1–9). In his often-cited study (1), Flynn related the skin permeability coefficients (K_p) of over 90 compounds to their octanol/water partition coefficients (K_{ow}). Potts and Guy (2) used multiple regression to develop an equation involving both K_{ow} and molecular weight (MW) to predict skin permeability coefficients and then tested that

algorithm with the Flynn database. In their revised model, Potts and Guy (3) related the skin permeability coefficient to molecular volume (MV) and hydrogen bond donor and acceptor activities by multiple regression analysis. Abraham *et al.* (6) also developed an equation using hydrogen bond activity parameters, which was very similar to Guy and Potts (3) but was able to predict skin permeability coefficients for additional compounds, including some steroids.

The purpose of this work was to use the above-mentioned, previously published databases of compounds to develop a fuzzy model to predict skin permeability coefficients. Fuzzy logic is a powerful tool that has been successfully used for modeling, control systems, pattern recognition, image processing and detection of distorted plethysmogram pulses (10–13). It differs from traditional Boolean logic in that input and output values to a fuzzy inference model can belong to multiple sets with different membership values in each set (14). Most human skin permeability models have been developed by postulating factors that may influence penetration and then using various mathematical techniques, typically multiple regression, to verify that influence. The objectives of this work were to assess the performance of a fuzzy inference system for predicting skin permeability coefficients and compare that to more traditional models.

In comparing these fuzzy models with previously published algorithms, the criteria for a "good" model are that its outputs closely correlate to experimental outputs (reflected in a correlation coefficient), it uses few inputs, it enhances understanding of the phenomenon, and it is easy to use.

THEORY

Introduction to Fuzzy Modeling

All modeling schemes, whether based on traditional mathematical principles or developed through fuzzy techniques, represent a mapping of a set of inputs to a set of outputs. For predicting chemical penetration through the skin, the output is the skin permeability coefficient and inputs include a variety of descriptors, such as MW, MV, log octanol/water partition coefficients, and hydrogen bonding activity (1–6). The difference between models previously published and the one described in this article is simply the method used to map the input to output. Many models are developed without a complete knowledge of the system being interpreted or predicted. For instance, analytical models are usually satisfactory at predicting outputs but can be oversimplifications of the system. This is because the system cannot be fully described, as all of the factors involved are either not known or completely understood. Independent of the method used, all traditional types of models impose a form of mapping based on known information. A set of conventions used to create a form or outline must be assumed to develop the model. An alternative method is to use a "model-free" form to map from a set of inputs to an output. In this case, natural rules are developed from the data rather than imposing rules on the modeling system. Therefore, in a model-free system, the rules are developed through clustering algorithms that divide the data into natural partitions. Mapping is then optimized through various techniques. The result of this "model-free" or

even "structure-free" model is still a mapping from inputs to outputs, similar to traditional algorithms.

The rules, whether imposed by the modeler or determined from the data, can be crisp (with the truth of the propositions being either true or false) or fuzzy (whereas the truths of propositions lie along a continuum along the unit interval). For instance, in Kirchner *et al.* (5), a chemical database is divided into subgroups of compounds with similar molar volumes. Linear regression is performed within each group to develop a relationship between log K_{ow} and log K_p. Through these subgroups or clusters (as described in the following section), Kirchner *et al.* (5) imposed rules on the system and the model was forced to conform to these rules. If, for instance, the data had clustered into subgroups, in which the members had partial membership or degree of belonging (as described in the following section), the resulting rules would be fuzzy.

Fuzzy Sets and Membership Functions

Phenomena to be modeled are complex and often riddled with uncertainty in the form of ambiguity. Traditionally, uncertainty is described in mathematical models by random characteristics, but fuzzy set theory allows this uncertainty to be represented through possibility rather than probability. Ross states that most uncertainty, however, is not truly random and can better be represented through fuzzy set theory (11).

Classic set theory defines objects as either a full or nonmember of a set. In fuzzy set theory, an element can be a partial member of the set. The element will have a degree of membership within a set, which can be defined by a particular membership function (14). For example, the set of chemicals with a MW from 500 to 700 is a classic or "crisp" set; the set of chemicals with MW in the region "around 600" is fuzzy. In a crisp or Boolean set, an element is either a full member of a particular set, represented by a membership value of one on the y-axis of Fig. 1a, or is not a member represented by a membership value of zero. In a fuzzy set, elements can have degrees of membership on the real continuous interval [0,1] with the endpoints of the interval (zero and one) again representing no or full membership, respectively, as demonstrated in Fig. 1b. Infinite values between these endpoints represent various degrees of membership for elements of the fuzzy set.

The use of membership functions is the major difference between crisp and fuzzy sets. In crisp sets, the membership function is unique, whereas fuzzy sets can have an unlimited number of membership functions and this flexibility translates to the utility of these types of sets (11). The membership function is the avenue through which to classify the fuzziness of a set. Membership functions can be assigned by intuition, inference, rank ordering, neural networks, genetic algorithms, and inductive reasoning, in addition to several other methods (11).

Clustering Methods

For complex systems or databases, an effective modeling approach can be to partition or "cluster" the available data into subsets and then approximate each subset by a simple model. By clustering data, not only can structures in the data
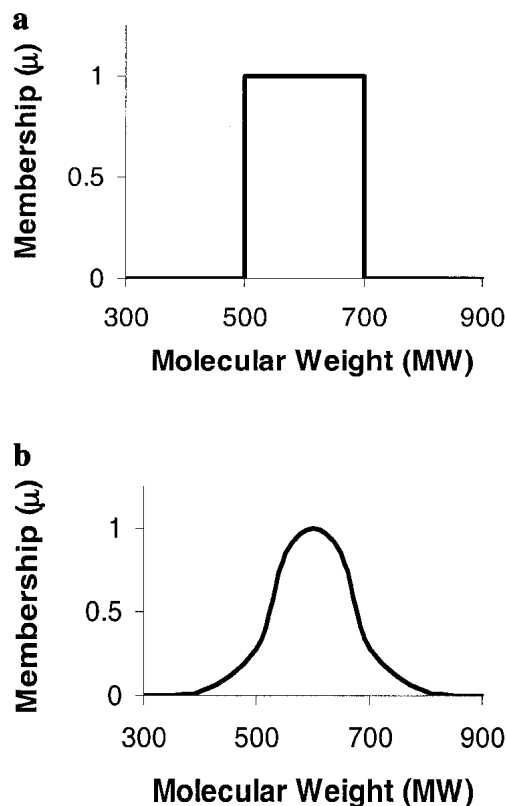

Fig. 1. Membership functions of (a) a crisp set and (b) a fuzzy set. The crisp set includes all compounds with a molecular weight (MW) of 500 to 700 and assigns them a membership value of one. All other compounds with an MW outside that set have membership equal to 0. The fuzzy set includes the set of compounds with an MW "around 600."

be revealed, but also the complexity of the model can be reduced. Hard or crisp clustering methods, based on classic set theory, partition the data into separate, mutually exclusive subsets and require complete or no membership of an object to a cluster. Fuzzy clustering methods allow objects to belong to multiple clusters at one time with different degrees of membership.

Fuzzy clustering is a means to organize data into groups. Each group is formed based upon similarity of the data belonging to that group or cluster. Data that belong to a particular cluster have more in common with other data belonging to that cluster than to data not belonging to that cluster. Further, given the nature of fuzzy sets, each data point can have membership in a variety of different clusters.

Procedurally, the degree of similarity, or membership, is determined in a mathematical sense (generally some distance norm). Particular aspects of these strategies depend upon the clustering method used. The range of complexity, appropriateness, and ease of implementation of fuzzy clustering methodologies is vast. A good treatment of the topic can be found in Babuška (12). Some of the fuzzy clustering methods include partitioning (15,16), subtractive clustering (17), fuzzy c-means, Gustafson-Kessel, fuzzy maximum likelihood estimate clustering, fuzzy c-varieties, fuzzy c-elliptotypes, fuzzy c-regression models, and possible clustering (12). Each of these methods has particular characteristics that dictate their respective usefulness for given situation. One of the limiting factors for these methods is that they are difficult to gener-

alize, and therefore it is difficult to develop the computational tools required for repetitive data analysis.

One of the most simple and straightforward clustering methods is fuzzy partitioning (16). This method divides each dimension of the data space into groups. The division is based entirely upon the scale of the data space without regard for the distribution or arrangement of the data within the data space. The resulting ordinate groups (or partitions) are the clusters that represent the input space. Similarly, the output data space is divided into groups. Rules are subsequently derived that map the data included in the input clusters to the clusters that represent the output space. This technique has merit and utility but is an unsophisticated approach to represent data and relationships between input and output data spaces.

Chiu (17) proposed subtractive clustering to mathematically evaluate each data point as a potential cluster center. After the initial examination, the point that has the greatest potential to be a cluster center is identified. Given that selection, the potential of every other data point to be a cluster center is reduced based upon its proximity to the identified cluster center. Subsequent points are chosen and this process continues until some stopping criteria are met. Given a variety of parameter designations, the result of this effort is a sense of data organization as expressed by the number of clusters identified, the location of the cluster centers, and the membership of each datum in each cluster. The procedures for subtractive clustering have been coded for use in the MatLab software (Version 6.1, Release 12.1, MathWorks, Inc., 2000) and this method is a good choice for modeling many forms of data. It is the method used in this work.

In a rule-based system, each cluster (describing the input space) developed by the above procedures has a corresponding rule (one rule per cluster). Each rule describes the relationship between the data in the cluster in the input space to the output space. Collectively, the compilation of rules maps the input data space to the output data space. For example, if the data space consisted of three variables, each of the rules would include three variables.

METHODS

The development of clusters and subsequent rules is conceptually sufficient to describe the mapping from the input space to the output space. However, refinement of the numeric parameters is useful to enhance the model fit. The means chosen was the Adaptive Neural Fuzzy Inference System (ANFIS; 16) as implemented in the MatLab Fuzzy Toolbox. A complete discourse on the implementation of these techniques is beyond the scope of this work. However, relevant features and context are described below.

The first step to using the fuzzy rule-based model is to develop the clusters and subsequent rules. The refinement of these rules is accomplished using by the Sugeno inference system (18), in which the antecedents are fuzzy propositions but consequences are crisp functions. The specific mapping to the crisp function is determined within ANFIS (16). The implementation of ANFIS using MatLab requires that the data set be divided into "training" and "checking" (sub)sets of the data. The training data (sub)set is representative of the entire data space. The selection of the training data (sub)set is accomplished in a variety of ways. In this work, one-half of the entire data set was selected as the training data. The checking data (sub)set is similarly selected.
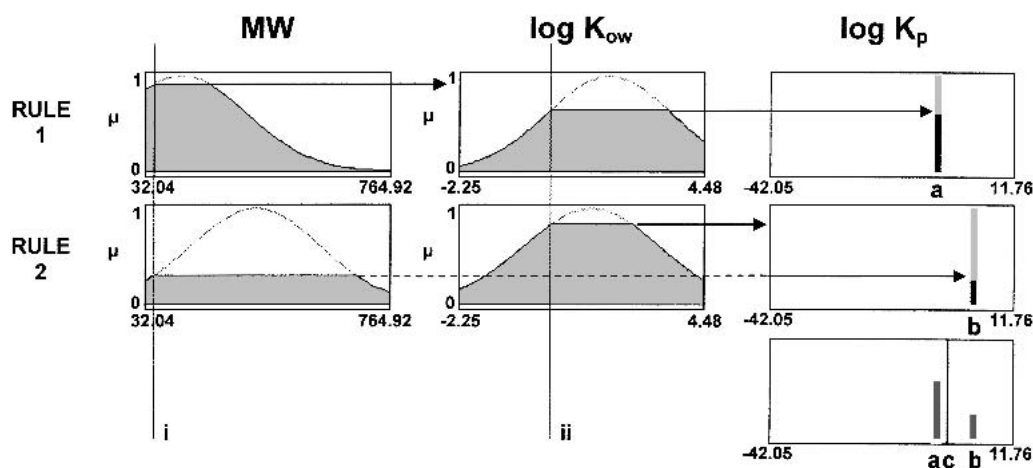

Fig. 2. MatLab (Mathworks, Inc., 2000) interactive interface describing Flynn Fuzzy Model. Each row of membership functions represents a rule and consists of two membership functions, one corresponding to each input. The first two columns represent the two inputs, molecular weight and log K_{ow}, respectively. The last column represents the weighted output (log K_p) of the each of the two rules. For each individual membership function, the range of input values is represented by the values on the x-axis (molecular weight or log K_{ow}) and the membership value (μ) is represented on the y-axis. The shaded region is a visual representation of the resulting membership value of the input value. The rule outputs (a and b) are represented by the multicolored bars. The gray portion represents the position of the rule output on the abscissa and the black portion of the bar represents the minimum of the two membership values of the inputs in each rule, indicated by the horizontal arrows. This interface gives access to the model, in which compound descriptors can be entered by moving the vertical lines (i and ii), which results in the weighted output (c). For a compound with molecular weight (i) and log K_{ow} (ii), the crisp output of the model (c) is the weighted average of the two crisp outputs (a and b) of the two rules.
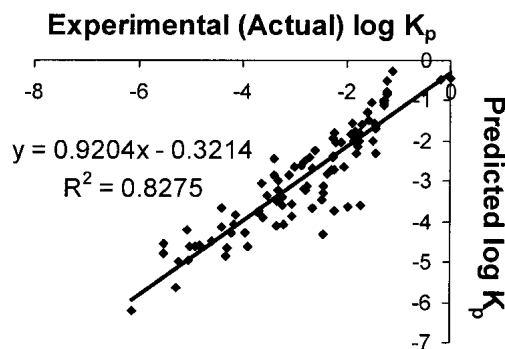
Experimental (Actual) log K_p

$y = 0.9204x - 0.3214$
$R^2 = 0.8275$

Fig. 3. Predicted skin permeability coefficients, as determined by Flynn Fuzzy Model, vs. experimental skin permeability coefficients from Flynn (1) with K_p in cm/h.

Using these data subsets, the ANFIS program modifies the membership functions of the antecedents to monitor a gradient vector that determines how well the fuzzy inference system is modeling the data set. Once this vector is determined, optimization of the parameters is accomplished to reduce errors, however the checking data (sub)set is used to monitor occurrences of over fitting.

Thus the skin permeability models presented here were developed by subtractive clustering and adaptive neural techniques that learned from the data obtained from previously published studies (1,3,6). Three separate models were developed using the ANFIS program in the MatLab software and each was based on a different group of inputs/outputs. The data were divided into two subsets, defined as the training and checking sets, which were used to train the model and then to prevent over fitting the data. The model was then evaluated by running the entire data set through it and this output data was then compared to the published experimental data.

Several fuzzy-ruled based models were developed to predict skin permeability coefficients of chemicals using various combinations of inputs. All fuzzy logic models predicting skin permeability were developed using MatLab software. The first model was developed using the database collected by Flynn (1), which includes MW and log octanol/water partition coefficients (log K_{ow}) for each chemical. These two descriptors served as the inputs to the model, and the log skin permeability coefficient (log K_p) was the output (as in all models). Next, models were developed with additional inputs. The second model was derived from the database compiled by Potts and Guy, (3) which is a subset of Flynn's original database and includes solvachromatic parameters for the compounds such as hydrogen bond donor activity (solute summation hydrogen-bond acidity, $\Sigma\alpha_2^H$), hydrogen bond acceptor activity (solute summation hydrogen-bond basicity, $\Sigma\beta_2^H$), dipolarity/polarizability (π), and the molar refractivity (R_2). The last fuzzy model to predict skin permeability coefficients was developed from the data compiled by Abraham *et al.*, (6) which was an extension of the Potts and Guy database, with the same inputs but some additional compounds.

Additional code was written to supplement the MatLab functions that enabled a combination of inputs to be tested in the model formulations. The "Flynn" fuzzy model (n = 94) resulted in a generation of three models for all combinations of MW and log K_{ow} as inputs. The Potts and Guy fuzzy model (n = 37) tested a combination of five inputs, including MV, $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$, π, as well as log K_{ow}, which was added as an input because it appeared from the Flynn model to be important. Molar refractivity (R_2) was not included as an input because it has been determined that it is not important (3). Finally, the Abraham fuzzy model data set (n = 54) expanded on the Guy and Potts database and used the same inputs as the Potts and Guy fuzzy model, with the replacement of MV with MW. By testing combinations of inputs, not only was the best fuzzy model determined, but the factors that are most important to predicting skin permeability were also discerned.
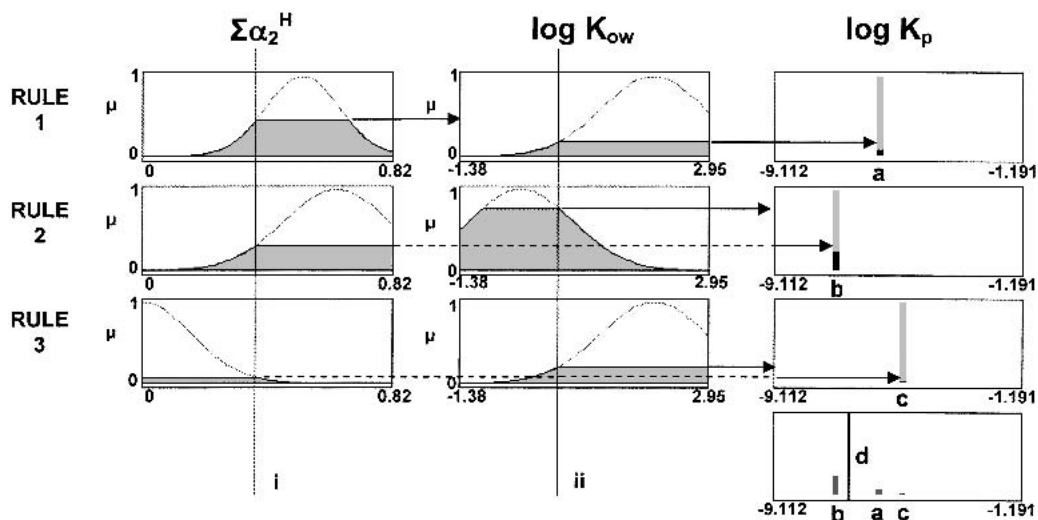

Fig. 4. MatLab (Mathworks, Inc., 2000) interface describing Potts and Guy Fuzzy Model. Each row of membership functions represents a rule. The first two columns represent the two inputs, $\Sigma\alpha_2^H$ and log K_{ow} respectively. The last column represents the weighted output (log K_p) of the each of the three rules. Range of input values is represented by the values on the x-axis and the membership value (μ) is represented on the y-axis. The shaded region is a visual representation of the resulting membership value of the input value. The rule outputs (a, b, and c) are represented by the multicolored bars. For a compound with $\Sigma\alpha_2^H$ (i) and log K_{ow} (ii), the crisp output of the model (d) is the weighted average of the crisp outputs (a, b, and c) of the three rules.
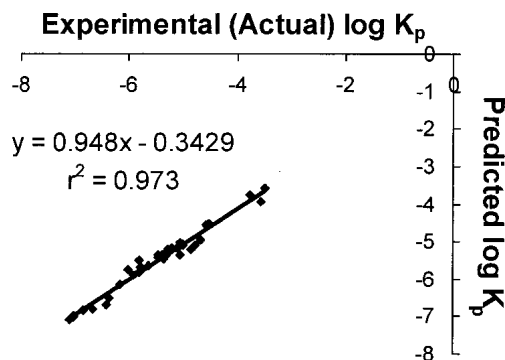
Fig. 5. Predicted skin permeability coefficients, as determined by Potts and Guy Fuzzy Model, vs. experimental skin permeability coefficients from Potts and Guy (3), with K_p in cm/s.

RESULTS AND DISCUSSION

For each of the three databases, all combinations of inputs were used to develop models, but only the best fuzzy model developed from each database is reported. The best Flynn fuzzy model included two inputs (MW and log K_{ow}) as was expected. The generated fuzzy model partitioned the data set (input/output) into two clusters and for each of these clusters, defined associated membership functions. For each cluster there was an associated rule as demonstrated in Fig. 2. Each row of membership functions represents a rule. The first two columns represent the two inputs, MW and log K_{ow}, respectively, and the last column represents the weighted output (log K_p) of the each of the two rules. The crisp output of the model is the weighted average of the two crisp outputs of the two rules. The weighting is based upon compliance (or membership) of the input to the cluster describing the input space. There was a good correlation ($r^2 = 0.82$) when the predicted output of this model was compared to experimental

values (Fig. 3). Previously published algorithms using multiple regression techniques (2) obtained a correlation coefficient of $r^2 = 0.67$ with the same two inputs. The results from the "Flynn" fuzzy model validate that MW and log K_{ow} can be used to predict skin permeability and demonstrates that fuzzy logic can successfully model this parameter as well as, if not better, than traditional analytical techniques (2).

The best Potts and Guy fuzzy model used a combination of just two inputs ($\Sigma\alpha_2^H$, termed H_d in Ref. 3 and log K_{ow}) and three clusters (Fig. 4). The additional third row (when compared with Fig. 2) represents the third cluster. Comparing the model outputs to experimental data resulted in an excellent correlation coefficient, $r^2 = 0.97$ (Fig. 5). A previous multiple regression model (3) using these hydrogen bonding activity factors, in addition to molar volume, resulted in a correlation coefficient of 0.94.

The best "Abraham" fuzzy model also used $\Sigma\alpha_2^H$ and log K_{ow} as inputs and evaluated these inputs using three rules (Fig. 6). When the outputs of these models were evaluated against experimental values (Fig. 7), an excellent correlation coefficient was observed ($r^2 = 0.95$) and corresponds to similar previous results using regression techniques (6). The advantage of the fuzzy model is the fewer number of inputs necessary to generate the output (two inputs in the fuzzy model vs. five inputs using regression techniques (6)).

Overall, the objective of these fuzzy models was to verify that fuzzy logic is a viable modeling alternative. To do this, a comparison between previously reported and similar fuzzy models was performed in Table I. From a comparison of several factors, including correlation coefficients and number of inputs required for successful mapping, it is obvious that fuzzy logic is a valid modeling approach that is at least as good as, if not better, than some other traditional methods. Additionally, this modeling system also reveals information about the system being modeled, as evidenced by the fact that the fuzzy
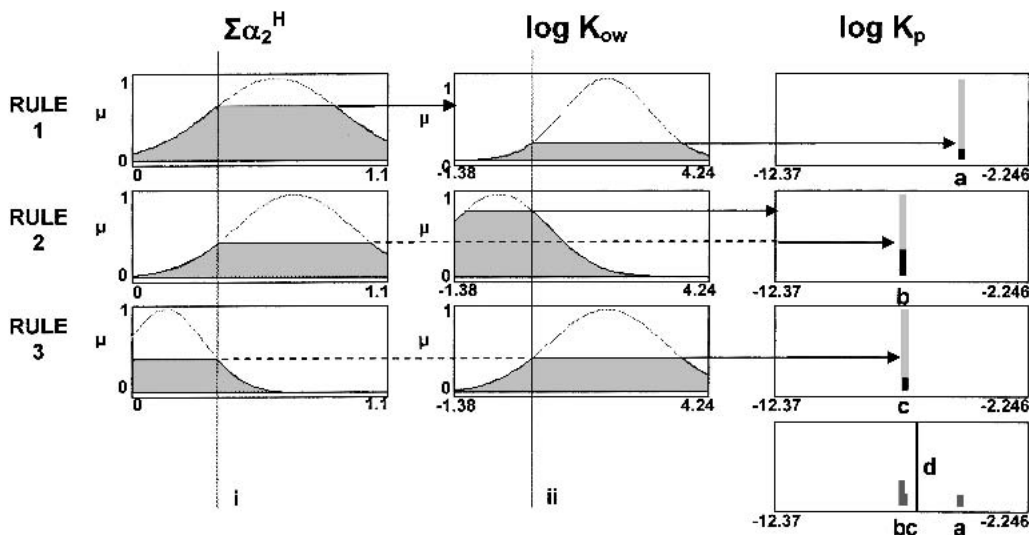

Fig. 6. MatLab (Mathworks, Inc., 2000) interface describing Abraham Fuzzy Model. Each row of membership functions represents a rule. The first two columns represent the two inputs, $\Sigma\alpha_2^H$ and log K_{ow} respectively. The last column represents the weighted output (log K_p) of the each of the three rules. Range of input values is represented by the values on the x-axis and the membership value (μ) is represented on the y-axis. The shaded region is a visual representation of the resulting membership value of the input value. The rule outputs (a, b, and c) are represented by the multicolored bars. For a compound with $\Sigma\alpha_2^H$ (i) and log K_{ow} (ii), the crisp output of the model (d) is the weighted average of the crisp outputs (a, b, and c) of the three rules.
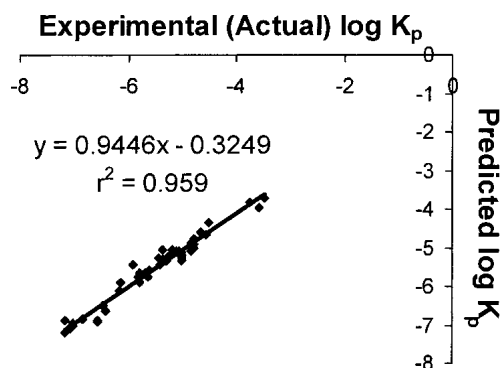
Experimental (Actual) log K$_p$

Fig. 7. Predicted skin permeability coefficients, as determined by Abraham Fuzzy Model, vs. experimental skin permeability coefficients from Abraham *et al.* (6), with K_p in cm/s.

models identified the same inputs as being central to predicting skin permeability, such as size and hydrogen bonding activity, which have previously been elicited by other modeling forms (1–8).

Fuzzy models were successful in predicting skin permeability coefficients. However, these models could be improved with different and perhaps more sophisticated clustering methods. The clustering methods available in the MatLab software are limited, but there have been other methods developed (16) not yet integrated into the software. Additionally, these fuzzy models were based on rather small databases. Increasing the number of compounds for which all input descriptors are available could create a data set that might better represent the data space and thus result in an improved model.

These models have proven that fuzzy logic is a realistic and promising tool that can successfully model skin permeability coefficients as well as or better than previous algorithms. More importantly, these fuzzy models reveal that fewer inputs are needed to predict skin permeability. Future studies should include attempts to study and extract informa-

tion from the models and cluster structures within, to gain a better understanding of the essential components of skin permeability. Additionally, future models will attempt to capitalize on the ability of fuzzy modeling to integrate expert opinion and conditional parameters. Thus with larger data sets, more sophisticated clustering techniques, and expert insight into the data structure, fuzzy logic could be the ideal approach to modeling skin permeability, along with many other parameters in medicine.

ACKNOWLEDGMENTS

REFERENCES

1. G. L. Flynn. Physicochemical determinants of skin absorption. In T. R. Gerrity and C. J. Henry (eds.), *Principles of Route-to-Route Extrapolation for Risk Assessment*, Elsevier, New York, 1990 pp. 93–127.
2. R. O. Potts and R. H. Guy. Predicting skin permeability. *Pharm. Res.* **9**:663–669 (1992).
3. R. O. Potts and R. H. Guy. A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity. *Pharm. Res.* **12**:1628–1633 (1995).
4. W. J. Pugh and J. Hadgraft. Ab initio prediction of human skin permeability coefficients. *Int. J. Pharm.* **103**:163–178 (1994).
5. L. A. Kirchner, R. P. Moody, E. Doyle, R. Bose, J. Jeffrey, and I. Chu. The prediction of skin permeability by using physicochemical data. *Altern. Lab. Anim.* **25**:359–370 (1997).
6. M. H. Abraham, F. Martins, and R. C. Mitchell. Algorithms for skin permeability using hydrogen bond descriptors: the problem of steroids. *J. Pharm. Pharmacol.* **49**:858–865 (1997).
7. T. Ghafourian and S. Fooladi. The effect of structural QSAR parameters on skin penetration. *Int. J. Pharm.* **217**:1–11 (2001).
8. P. Poulin and K. Krishnan. Molecular structure-based prediction of human abdominal skin permeability coefficients for several organic compounds. *J. Toxicol. Environ. Health* **62**:143–159 (2001).
9. D. A. Edwards and R. Langer. A linear theory of transdermal transport phenomena. *J. Pharm. Sci.* **83**:1315–1334 (1994).
10. H. R. Tizhoosh. Fuzzy image enhancement: an overview. In E. E. Kerre and M. Nachtegael (eds.), *Fuzzy Techniques in Image Processing*, Physica-Verlag, New York, 2000 pp. 136–171.
11. T. J. Ross. *Fuzzy Logic with Engineering Applications*, McGraw-Hill, Inc., New York, 1995.
12. R. Babuška. *Fuzzy Modeling for Control*, Kluwer Academic Publishers, Boston, Massachusetts, 1998.
13. S. Y. Belal, A. F. G. Taktak, A. J. Nevill, S. A. Spencer, and D. Roden, S. Bevan. Automatic detection of distorted plethysmogram pulses in neonates and paediatric patients using an adaptive-network-based fuzzy inference system. *Artif. Intell. Med.* **4**: 149–165 (2002).
14. L. A. Zadeh. Fuzzy sets. *Inf. Control* **8**:338–353 (1965).
15. *Fuzzy Logic Toolbox User's Guide. Version 2*, The MathWorks, Inc., Natick, Massachusetts, 2000.
16. J. S. R. Jang. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **23**:665–684 (1993).
17. S. L Chiu. Fuzzy model identification based on cluster estimation. *J. Intell. Fuzzy Syst.* **2**:267–278 (1994).
18. M. Sugeno. *Industrial Applications of Fuzzy Control,* Elsevier Science Publishing Co., New York, 1985.

Table I. Comparison of Previously Published Models to Three Fuzzy Models

Model reference	Database reference	Inputs	n	r^2
(2)[a]	(1)	log K_{ow}, MW	93	0.670
(3)	(3)	MV, $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$	37	0.940
(6)	(6)	R_2, π, V_x, $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$	53	0.958
A	(1)	log K_{ow}, MW	93	0.828
B	(3)	$\Sigma\alpha_2^H$, log K_{ow}	37	0.973
C	(6)	MW, $\Sigma\alpha_2^H$	53	0.959

A. Fuzzy model described in Figs. 2 and 3 (Flynn Fuzzy Model).
B. Fuzzy model described in Figs. 4 and 5 (Potts and Guy Fuzzy Model).
C. Fuzzy model described in Figs. 6 and 7 (Abraham Fuzzy Model).
MW, molecular weight.
[a] Refers to reference numbers.