# Quantum Pharmacological Analysis of Structure−Activity Relationships for Mefloquine Antimalarial Drugs Using Optimal Transformations

Vu Nguyen-Cong* and Bernd M. Rode*

Institute for General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innrain 52a, A-6020 Innsbruck, Austria

Received June 23, 1995[⊗]

Optimal nonlinear transformations provided by alternating conditional expectations (ACE) method are used to study the relationship between electronic structure and antimalarial activity of mefloquine and its substituted derivatives. The electronic structure of these molecules is featured by atomic net charges evaluated on the basis of CNDO/2 molecular orbital calculations. A comparison to multiple linear regression (MLR) and partial least squares (PLS) method is also included. The results show that ACE can be a useful and well-suited technique for establishing QSAR and predicting the pharmacological activity of compounds.

## INTRODUCTION

Mefloquine [*dl*-erythro-α-(2-piperidyl)-2,8-bis(trifluoromethyl)-4-quinoline methanol] (Figure 1), originally synthesized by Lutz et al.[1], is a blood schizontocide active against plasmodia in the intraerythrocytic stage of their life cycle. Like many other antimalarial agents, the eventual emergence of resistance to mefloquine is inevitable. However, with cautious use and possible combination with other agents, such as with artemisinin derviatives,[2,3] mefloquine probably retains its efficiency in the treatment and prophylaxis of malaria tropica, one of the most serious diseases in developing countries worldwide.

During the past 30 years, hundreds of different descriptors describing structural features have been used in quantitative structure−activity relationships (QSAR). These parameters have been recently discussed by Kubinyi,[4] Jurs et al.,[5] and Kier.[6] However, very little has been published in the chemical literature concerning the relationship between molecular electronic structure derived from quantum mechanical calculations and pharmacological activity. A previous study[7] has shown a linear relationship between antimalarial activity of mefloquine-type compounds and atomic net charges based on semiempirical CNDO/2 MO SCF calculations. The purpose of this work is to build nonlinear predictive models using a nonparametric regression technique called alternating conditional expectations[8] (ACE) and to compare results achieved by the best ACE model to the best ones obtainable from multiple linear regression[9] (MLR) and partial least squares[10−12] (PLS) methods. The quality of a model is rated high if it uses as few descriptors as possible but still provides good predictive ability.

## METHOD

**Alternating Conditional Expectations (ACE).** ACE is an additive regression model introduced by Breiman and Friedman.[8] It has the following form

$$\theta(y) = \phi_1(x_1) + \phi_2(x_2) + ... + \phi_p(x_p) + \mathrm{r}$$

where $y$ is the response, each of $x_i$ are predictors, $\theta$ and $\phi_i$ are nonlinear transformations with expectation zero $E\theta(y) = E\phi_i(x_i) = 0$, $E\theta^2(y) = 1$, and $r$ is the residual. ACE applies an iterative procedure for finding optimal transformations that minimize the mean squared error

$$e^2 = \frac{1}{n}\sum_{i=1}^{n}[\theta(y_i) - \sum_{j=1}^{p}\phi_j(x_{ij})]^2$$

where $n$ is the number of observations. In other words, ACE provides nonlinear transformations of both the predictors and the response to maximize the correlation between the transformed response and the sum of the transformed predictors. The multiple $R$-squared value for the transformed values is given by $R^2 = 1 - e^2$. The ACE transformations are obtained using a two-dimensional scatterplot smoother called *supersmoother*.[13,14] Supersmoother is a variable span smoother based on local linear fits, with the optimal span chosen by local cross-validation.

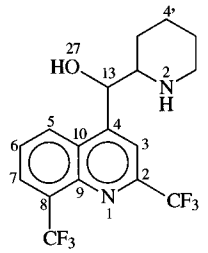


**Figure 1.** Structure of mefloquine.

**Cross-Validation.** A major goal in QSAR studies is to predict biological responses of new compounds using the model derived from investigated compounds. Cross-validation[15] (CV) is a tool which is nearly always used for this purpose. CV proceeds by leaving out one or more compounds randomly from the dataset, rebuilding the model, and predicting the response for the compounds left out. The process is repeated a number of times so that each compound is left out exactly once. Then the sum of squares of the residuals between actual and predicted response values is calculated, giving the so-called *PRESS* statistic (predictive residual sum of squares) value.

This work uses the leave-one-out approach for CV. In calculating the PRESS statistic, for a given set of $p$ predictors, each observation $y_i$, is predicted from MLR, PLS, or ACE models obtained from the other $(n - 1)$ observations. If $\hat{y}_{ip}$

[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1995.

0095-2338/96/1636-0114$12.00/0 

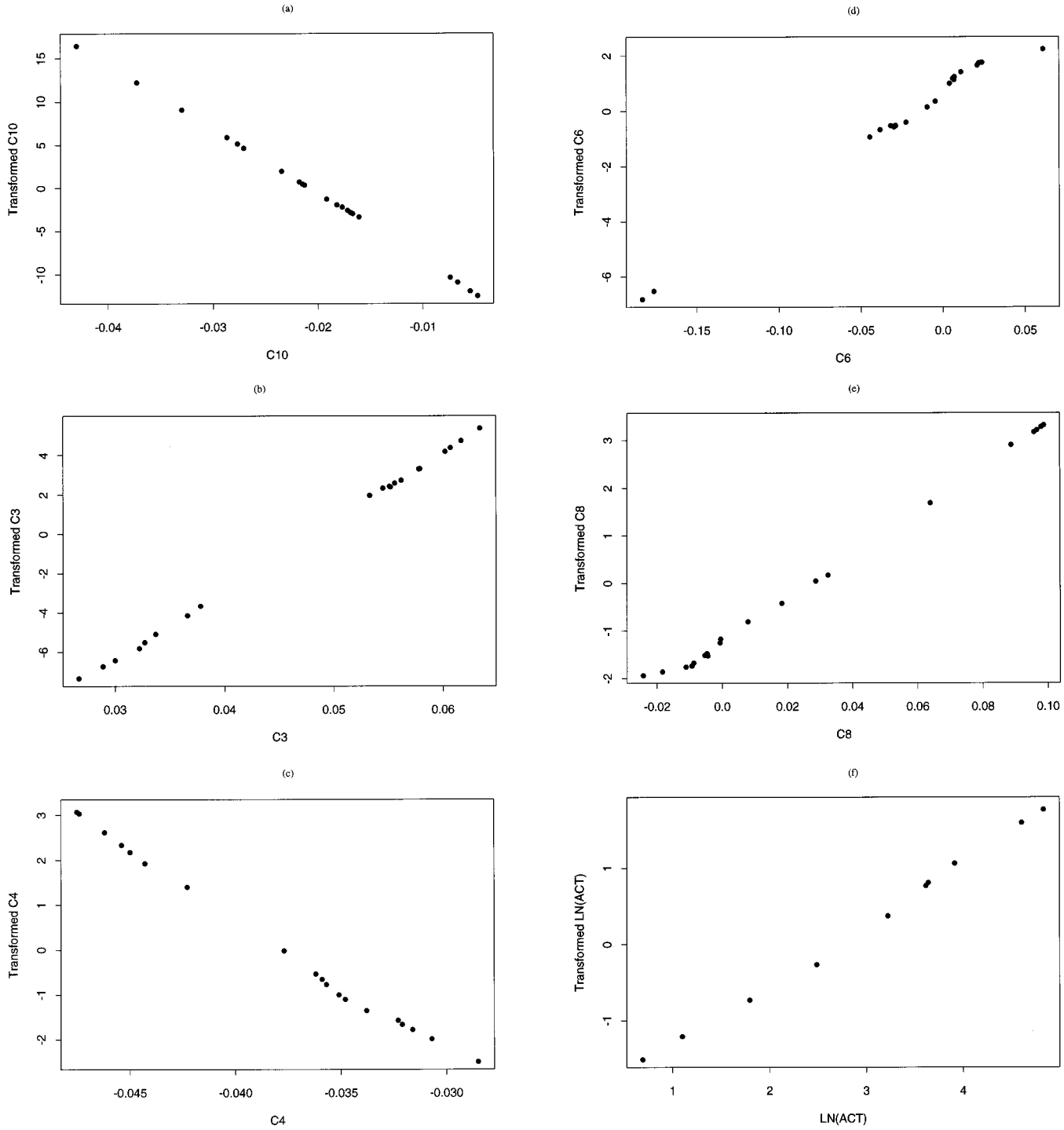


**Figure 2.** Plots of ACE fit of the mefloquine dataset.

denotes the predicted value for $y_i$, the *PRESS* statistic for a particular subset of $p$ predictors is

$$PRESS_p = \sum_{i=1}^{n} (y_i - \hat{y}_{ip})^2$$

The predictive performance of a QSAR model can also be described by the cross-validated $r^2$ value.[16] It is defined as

$$r^2_{cv} = 1.0 - \frac{\text{PRESS}}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

where $\bar{y}$ is the mean of the responses. A $r^2_{cv}$ value of 1.0 corresponds to perfect predictions.

Our program to perform ACE algorithm is based on the Friedman's FORTRAN subroutines stored in StatLib archive with a few minor modifications. ACE calculations were performed at an IBM/RS6000 workstation and a Sybyl Programming Language (SPL) program performing PLS at a Silicon Graphics workstation of the Institute for General, Inorganic and Theoretical Chemistry. A *S* language[17,18] program for MLR calculations was run on a Silicon Graphics Workstation of the Institute for Statistics, University of Innsbruck.

The data for this study were taken from a previous work.[7] It contains antimalarial activity values in mice of 21

**Table 1.** Chemical Structures and Antimalarial Activities of Mefloquine Derivatives Relative to Mefloquine (=100%)

| compd no. | nuclear substituent (X) | antimalarial activity, % |
|---|---|---|
| MF1 | 8-$CF_3$, 2-$CF_3$ | 100 |
| MF2 | 6-$OCH_3$, 8-$CF_3$, 2-$CF_3$ | 25 |
| MF3 | 7-$CF_3$, 2-$CF_3$ | 38 |
| MF4 | 6-$CF_3$, 2-$CF_3$ | 12 |
| MF5 | 6-$OCH_3$, 2-$CF_3$ | 6 |
| MF6 | 6-$CH_3$, 2-$CF_3$ | 2 |
| MF7 | 8-$CH_3$, 2-$CF_3$ | 2 |
| MF8 | 6-$CH_3$, 8-$CH_3$, 2-$CF_3$ | 3 |
| MF9 | 6-$CH_3$, 8-$CH_3$, 4′-$CH_3$ | 3 |
| MF10 | 6-$CH_3$, 8-$CH_3$, 4′-$OCH_3$ | 12 |
| MF11 | 6-$CH_3$, 8-$CH_3$, 4′-Cl | 38 |
| MF12 | 6-$CH_3$, 8-$CH_3$, 4′-F | 12 |
| MF13 | 8-$CF_3$, 4′-H | 50 |
| MF14 | 8-$CF_3$, 4′-$CH_3$ | 25 |
| MF15 | 8-$CF_3$, 4′-$OCH_3$ | 37 |
| MF16 | 8-$CF_3$, 4′-Cl | 125 |
| MF17 | 6-$CH_3$, 4′-$OCH_3$ | 3 |
| MF18 | 8-$HC_3$, 4′-H | 6 |
| MF19 | 8-$CH_3$, 4′-$CH_3$ | 6 |
| MF20 | 8-$CH_3$, 4′-Cl | 50 |
| MF21 | 8-$CH_3$, 4′-F | 25 |

mefloquine derivatives (Table 1) and 13 net atomic charges at the molecular positions denoted in Figure 1. The response was converted into natural logarithmic scale before doing QSAR analyses.

## RESULTS AND DISCUSSION

The basic approach was to investigate all subsets of the 13 electronic descriptors and to build models with high values of both $R^2$ and $r^2_{cv}$. Although the number of degrees of freedom is low in the beginning of analysis, 21 compounds versus 13 variables, all three methods give an optimal $r^2_{cv}$ with the same set of five predictors (C3, C4, C6, C8, and C10). Both MLR and PLS produce identical values of coefficients and $R^2$ for the final five-descriptor model. Tables 2 and 3 gives fitting and cross-validation data of best MLR/PLS and ACE models, respectively. They show that ACE achieves not only a better fit but also a better predictive power than MLR/PLS. However, it is surprising that the parent compound (MF1) was so poorly predicted, and the data used in this analysis did not give an apparent reason for it. Figure 2a−f shows original versus transformed values for the response and predictors appearing in the final model. The plots indicate that the relations between original and transformed values of the predictors are nearly linear. This explains why the predictive power of the ACE model is not much higher than that of the MLR/PLS model. However, a recent study[19] on antibiotic drugs of cephalosporin type has shown that the ACE technique can outperform the MLR method, whenever a nonlinear QSAR exists.

Mechanisms for antimalarial action of quinoline-containing compounds have been recently reviewed by Palmer et al.[20] Karle and Karle[21] have suggested that the orientation of the hydroxyl and amine groups with respect to each other may be crucial for antimalarial activity because of the hydrogen bonding ability of these functional units. However, our final models based on the 21-observation dataset here contain only carbon atoms of the quinoline ring, thus emphasizing the role of hydrophobic interaction.

Concerning the influence of predictors in a QSAR model, the SYBYL/QSAR PLS uses normalized regression coefficients[22] whilst standard deviations of transformed predictors are used in ACE. Table 4 shows that the MLR/PLS and ACE models are in agreement with each other in providing the measure of how strongly each predictor contributes to the resulting QSAR model.

**Table 2.** Details of the Best MLR/PLS Model

| compd no. | obsd ln(ACT) | fitted ln(ACT) | residual (obsd − fit) | predicted (leave-1-out) | residual (obsd − pred) |
|---|---|---|---|---|---|
| MF1 | 4.605 | 3.432 | 1.173 | 2.951 | 1.654 |
| MF2 | 3.219 | 3.684 | −0.465 | 4.282 | −1.063 |
| MF3 | 3.638 | 3.964 | −0.326 | 4.177 | −0.539 |
| MF4 | 2.485 | 3.216 | −0.731 | 3.656 | −1.171 |
| MF5 | 1.792 | 1.458 | 0.334 | 1.123 | 0.669 |
| MF6 | 0.693 | 1.125 | −0.432 | 1.541 | −0.848 |
| MF7 | 0.693 | 0.786 | −0.093 | 0.848 | −0.155 |
| MF8 | 1.099 | 0.560 | 0.539 | 0.348 | 0.751 |
| MF9 | 1.099 | 1.621 | −0.522 | 1.719 | −0.620 |
| MF10 | 2.485 | 2.251 | 0.234 | 2.218 | 0.267 |
| MF11 | 3.638 | 3.125 | 0.513 | 2.914 | 0.724 |
| MF12 | 2.485 | 2.661 | −0.176 | 2.691 | −0.206 |
| MF13 | 3.912 | 3.753 | 0.159 | 3.714 | 0.198 |
| MF14 | 3.219 | 3.154 | 0.065 | 3.131 | 0.088 |
| MF15 | 3.611 | 3.966 | −0.355 | 4.112 | −0.501 |
| MF16 | 4.828 | 4.755 | 0.073 | 4.731 | 0.097 |
| MF17 | 1.099 | 1.747 | −0.648 | 1.921 | −0.822 |
| MF18 | 1.792 | 1.897 | −0.105 | 1.913 | −0.121 |
| MF19 | 1.792 | 1.846 | −0.054 | 1.856 | −0.064 |
| MF20 | 3.912 | 3.905 | 0.007 | 3.903 | 0.009 |
| MF21 | 3.219 | 2.404 | 0.815 | 2.199 | 1.020 |
| | standard error of estimate | | | | 0.560 |
| | multiple r-squared | | | | 0.856 |
| | F-statistic | | | | 17.862 |
| | *PRESS* | | | | 10.347 |
| | cross-validated *r*-squared | | | | 0.684 |

**Table 3.** Details of the Best ACE Model

| compd no. | obsd ln(ACT) | fitted ln (ACT) | residual (obsd − fit) | predicted (leave-1-out) | residual (obsd − pred) |
|---|---|---|---|---|---|
| MF1 | 4.605 | 4.449 | 0.156 | 3.014 | 1.591 |
| MF2 | 3.219 | 3.465 | −0.246 | 3.950 | −0.731 |
| MF3 | 3.638 | 3.834 | −0.196 | 4.159 | −0.521 |
| MF4 | 2.485 | 2.578 | −0.093 | 1.865 | 0.620 |
| MF5 | 1.792 | 1.453 | 0.339 | 0.618 | 1.174 |
| MF6 | 0.693 | 0.627 | 0.066 | 1.243 | −0.550 |
| MF7 | 0.693 | 0.794 | −0.101 | 1.029 | −0.336 |
| MF8 | 1.099 | 0.990 | 0.109 | 0.777 | 0.322 |
| MF9 | 1.099 | 1.550 | −0.451 | 1.951 | −0.852 |
| MF10 | 2.485 | 2.240 | 0.245 | 1.823 | 0.662 |
| MF11 | 3.638 | 3.350 | 0.288 | 2.993 | 0.645 |
| MF12 | 2.485 | 2.399 | 0.086 | 2.647 | −0.162 |
| MF13 | 3.912 | 3.713 | 0.199 | 3.733 | 0.179 |
| MF14 | 3.219 | 3.280 | −0.061 | 3.238 | −0.019 |
| MF15 | 3.611 | 4.010 | 0.399 | 4.111 | −0.500 |
| MF16 | 4.828 | 4.798 | 0.030 | 4.208 | 0.620 |
| MF17 | 1.099 | 1.192 | −0.093 | 1.993 | −0.894 |
| MF18 | 1.792 | 2.055 | −0.263 | 2.095 | −0.303 |
| MF19 | 1.792 | 1.819 | −0.027 | 1.972 | −0.180 |
| MF20 | 3.912 | 3.781 | 0.131 | 3.748 | 0.164 |
| MF21 | 3.219 | 2.876 | 0.343 | 2.221 | 0.998 |
| | multiple *r*-squared | | | | 0.961 |
| | *PRESS* | | | | 9.841 |
| | cross-validated *r*-squared | | | | 0.699 |

**Table 4.** Contributions of Predictors to the PLS and ACE Models

| variable name | norm coeff (MLR/PLS) | std dev (ACE) |
|---|---|---|
| C10 | 8.265 | 7.426 |
| C3 | 4.844 | 4.425 |
| C4 | 1.666 | 1.812 |
| C6 | 2.763 | 2.347 |
| C8 | 2.296 | 1.985 |

In conclusion, application of the optimal transformations on semiempirical electronic data of biological compounds represents a valuable tool besides linear regression techniques for explaining and predicting pharmacological activity of compounds. Since the ACE technique overcomes the restrictions of linear models, it seems to be a more general and versatile tool for all works of QSAR analyses.

## ACKNOWLEDGMENT

A grant from the Austrian Federal Government for Vu Nguyen-Cong is gratefully acknowledged. We would also like to thank Dr. Romano Kroemer, Sandoz RS, for a very helpful conversation on SPL programming.

## REFERENCES AND NOTES

(1) Ohnmacht, C. J.; Patel, A. R.; Lutz, R. E. Antimalarials. 7. Bis-(trifluoromethyl)-α-(2-piperidyl)-4-quinolinemethanols. *J. Med. Chem.* **1971**, *14*, 926−928.

(2) Looareesuwan, S. Overview of Clinical Studies on Artemisinin Derivatives in Thailand. *Trans. R. Soc. Trop. Med. Hy.* **1994**, *88 Suppl 1*, S9-11.

(3) Arnold, K. Early Treatment of Malaria in the Community using Artemisinin−Hope or Hazard? *Trans. R. Soc. Trop. Med. Hy.* **1994**, *88 Suppl 1*, S47−49.

(4) Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*. VCH: Weinheim, 1993; Chapter 3, p 21.

(5) Jurs, P. C.; Dixon, S. L.; Egolf, L. M. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995; Vol. 2, Chapter 2, p 15.

(6) Kier, L. B. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995; Vol. 2, Chapter 2, p 39.

(7) Polman, S.; Kokpol, S.; Hannongbua, S.; Robe, B. M. Quantum Pharmacological Analysis of Structure-Activity Relationships for Mefloquine Antimalarial Drugs. *Analyt. Sci.* **1989**, *5*, 641−644.

(8) Breiman, L.; Friedman, J. H. Estimating Optimal Transformations for Multiple Regression and Correlation (with discussion). *J. Amer. Statist. Assoc.* **1985**, *80*, 580−619.

(9) Weisberg, S. *Applied Linear Regression*; Second Edition; John Wiley & Sons: New York, 1985.

(10) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. III The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735−743.

(11) Hellberg, S.; Sjöström, M.; Skagerberg, B.; Wold, S. Peptide Quantitative Structure−Activity Relationships, a Multivariate Approach. *J. Med. Chem.* **1987**, *30*, 1126−1135.

(12) Wold, S. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995; Vol. 2, Chapter 4, p 195.

(13) Friedman, J. H.; Stuetzle, W. Smoothing of Scatterplots. *Technical Report ORION006*, Stanford University, Dept. of Statistics, 1982.

(14) Friedman, J. H. A Variable Span Smoother. *Technical Report LCS05*, Stanford University, Dept. of Statistics, 1984.

(15) Stone, M. Cross-validatory Choice and Assessment of Statistical Predictions (with discussion). *J. Roy. Stat. Soc. Ser. B* **1974**, *36*, 111−147.

(16) Cramer, R. D. III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(17) Becker, R. A.; Chambers, J. M.; Wilks, A. R. *The NEW S Language*; Chapman and Hall: New York, 1988.

(18) Chambers, J. M. In *Statistical Models in S*; Chambers, J. M.; Hastie, T. J., Ed.; Chapman and Hall: New York, 1992; Chapter 4, p 95.

(19) Nguyen-Cong, V.; Rode, B. M. Application of Nonparametric Regression Methods to Quantitative Electronic Structure-Activity Relationship (QSAR) Study for Antibiotic Drugs of Cephalosporin Type. 10th International Worshop on Statistical Modelling, Innsbruck, Austria, 1995.

(20) Palmer, K. J.; Holliday, S. M.; Brogden, R. N. Mefloquine. A Review of its Antimalarial Activity, Pharmacokinetic Properties and Therapeutic Efficacy. *Drugs* **1993**, *3*, 430−475.

(21) Karle, J. M.; Karle, I. L. Crystal Structure and Molecular Structure of Mefloquine Methylsulfonate Monohydrate: Implications for a Malaria Receptor. *Antimicrob. Agents Chemother.* **1991**, *35*, 2238−2245.

(22) SYBYL. Molecular Modeling Software. Theory Manual: TRIPOS Associates, Inc.: p 2256.

CI9502459