Microcomputer-Based Nonlinear Regression Analysis of Ligand-Binding Data: Application of Akaike’s Information Criterion

Keita KAMIKUBO, Hiroshi MURASE, Masanori MURAYAMA and Kiyoshi MIURA

The Third Department of Internal Medicine, Gifu University School of Medicine, 40 Tsukasa-machi, Gifu 500, Japan

Accepted November 18, 1985

Abstract—Akaike’s information criterion (AIC) (Akaike, H., IEEE Trans. Automat. Contr. AC-19, 716–723 (1974)) was applied to estimate statistically the number of classes of binding sites from ligand-binding data. Several sets of data were analyzed by both the AIC method and the F-test method. Good agreement was obtained between results from both methods. The present results suggest that the AIC method can be a good alternative to the F-test to estimate the number of classes of sites.

Because of rapidly increasing numbers of ligand-binding studies such as radioligand/receptor binding studies, it has become very important to analyze ligand-binding data accurately. It has been well documented that graphical or linear least-squares analyses of transformed ligand-binding data such as Scatchard analysis do not necessarily provide accurate estimates of binding parameters (1–4). In addition, it is not easy to analyze concave-up Scatchard plots by graphical methods (5). On the other hand, the computerized, nonlinear least-squares analysis of untransformed data (Bound vs. Free) can easily provide optimal estimates of parameters (2, 6, 7).

When analyzing ligand-binding data by the nonlinear least-squares procedure, the site heterogeneity model is often employed because of its simplicity. In the site heterogeneity model, an observed binding isotherm is regarded as a summation of bindings to several classes of binding sites. The most critical problem in analyzing ligand-binding data with the site heterogeneity model is estimation of the number of classes of specific binding sites. In general, resolving binding data into a model with a larger number of classes of binding sites improves the fitness of the regression curve to experimental data. However, because the information carried by observed data is limited, uncertainty of estimated parameters increases as a model of too high order is selected. Without knowledge from other types of experiments or an a priori assumption, therefore, the number of classes of binding sites should be determined by an adequate statistical procedure.

The statistical estimation of the number of classes of binding sites has been exclusively performed by applying a partial F-test for the residual sum of squares in the nonlinear regression analysis (2, 7, 8). However, it seems that little effort has been made to simplify the method for estimating the number of classes of sites. In the present study, we have applied Akaike’s information criterion (AIC) (9) to select the number of classes of binding sites statistically. AIC numerically expresses “the unlikelihood” of a model. Selecting an order of model that gives the minimum AIC is asymptotically equivalent to selecting an order of model with minimum Kullback-Leibler’s mean information (9, 10), i.e., an order of model with the maximum likelihood.

Ligand-binding data from several real experiments and computer simulations were
analyzed by the weighted nonlinear least-squares method. Binding data were obtained as follows. 1) Binding of \([3^H]\)diprenorphine (41 Ci/mmol, Amersham) and \([125^I]\)\(\beta\)-endorphin (1870 Ci/mmol, Amersham) to bovine adrenal medullary membranes (plasma membrane-mitochondrial fraction, prepared as described by Wilson and Kirshner (11)), or rat brain synaptosomal membranes (prepared as described elsewhere (12)) was assayed essentially as described previously (13). Briefly, these membranes were incubated with increasing amounts of each radioligand in 0.05 M Tris-HCl (pH 7.4) at 25° C for 30 min. For the assay of \([125^I]\)\(\beta\)-endorphin binding, 0.1% bovine serum albumin and 0.05% bicitracin were included in the incubation mixture. Nonspecific binding was determined in the presence of 1 \(\mu\)M of each unlabeled ligand. Bound radioligand was separated from free ligand by rapid vacuum filtration through Whatman GF/C filters. 2) Artificial ligand-binding data were generated by adding normally distributed errors to the amount of bound ligand. The amount of bound ligand (B) was calculated according to the multiple equilibrium model based on the mass action law (6) as: 

\[
B = (1 + \varepsilon) \cdot \frac{F - B_{\text{max},i}}{F + K_{d,i}}
\]

where F is the concentration of free ligand, \(K_{d,i}\) and \(B_{\text{max},i}\) are the dissociation constant and the maximum binding capacity of the i-th class of binding sites, and \(\varepsilon\) is an error with mean of zero and fixed variance.

Each set of binding data (Bound vs. Free) was analyzed by the weighted nonlinear least-squares procedure using the Fletcher-Marquardt-Gauss-Newton algorithm (14). The computer program written in BASIC (MULTI) (15) was used for the nonlinear curve fitting. The multiple equilibrium model based on the mass action law was used in the program.

AIC was calculated as:

\[
AIC = N \cdot \ln S + 2P
\]

where N and P denote the number of data and the number of parameters to be estimated, respectively; \(S\) denotes the residual sum of squares (9, 10). The number of parameters for a single class of specific binding sites is two (\(K_d\) and \(B_{\text{max},i}\)), and that for nonspecific binding is one (fractional binding). A model with the number of parameters that gave a minimum AIC was selected as the most likely one.

The F-test was performed for each pair of number of parameters. The F-value was calculated as:

\[
F = \frac{(S_1 - S_2) / (df_1 - df_2)}{S_2 / df_2}
\]

where \(S_1\) and \(S_2\) are the residual sums of squares for one and another model, respectively; \(df_1\) and \(df_2\) are the corresponding degrees of freedom (number of data minus number of parameters) (2, 7). The significance limit of each test was set at 95%.

Calculations were performed by using the program written in double precision BASIC for a FACOM 9450 microcomputer.

Figure 1A graphically shows nonlinear least-squares analysis of specific \([3^H]\) diprenorphine binding to bovine adrenal medullary membranes. The residual sum of squares decreased from 8.43 to 8.23 by increasing the number of parameters of a model from two to three—from a model of a single class of specific sites to a model of a single class of specific sites plus nonspecific sites. This improvement of fitness, however, was not significant as judged by the F-test at 95% significance limit. Thus, the F-test method selected the model of a single class of specific binding sites. AIC increased from 33.8 to 35.5 by increasing the number of parameters from two to three. So, the AIC method also selected the same model as the F-test did.

Figure 1B shows results of similar analysis of total \([125^I]\)\(\beta\)-endorphin binding to rat brain synaptosomal membranes. Residual sums of squares were 71.2 and 0.55, and the corresponding values of AIC were 38.1 and 1.19 for the number of parameters of two and three. The improvement of fitness by increasing the number of parameters from two to three was significant as judged by the F-test. A fitting to a model with four parameters led to extremely large absolute values (>10^18) of both \(K_d\) and \(B_{\text{max},i}\) for the second class of specific binding sites. The residual sum of squares for the model with four parameters was indistinguishable from that for the model with three parameters. So,
the value of AIC for the model with four parameters was 3.19. The model with three parameters, a single class of specific binding sites plus nonspecific binding sites, was, therefore, selected by both the AIC method and the F-test method. From a physical point of view, the vast values of the binding parameters for the putative second class of sites means that the second sites are virtually nonspecific. These considerations also support the model selected by the statistical methods.

Analysis of specific [125I]β-endorphin binding to bovine adrenal medullary membranes is shown in Fig. 1C. The residual sums of squares were 57.0, 12.4 and 1.73 for the number of parameters of two, three and four, respectively. A fitting to a model with five parameters did not converge. The F-test method selected the number of parameters as four—two classes of specific sites. Values of AIC were 36.3, 26.1 and 12.4 for the number of parameters of two, three and four, respectively. In all this case, therefore, both the AIC method and the F-test method selected the same model of binding—two classes of specific binding sites.

In order to study performance of the AIC method further, ligand-binding data were artificially generated and analyzed by the nonlinear regression procedure using either the F-test method or the AIC method. Three representative sets of data were generated based on a model of a single class of specific sites, a single class of specific sites plus nonspecific sites or two classes of specific sites. In all these cases, as shown in Fig. 2, both the AIC method and the F-test method
selected the correct models.

One of the most important problems in analyzing ligand-binding data by the site heterogeneity model is determination of the number of classes of specific binding sites. The determination would be troublesome without an *a priori* assumption or knowledge from other types of experiments. Between the goodness of fit and the uncertainty of estimated parameters, an appropriate decision-making procedure is required to determine the number of classes of sites under these situations. We believe that a statistical approach is one of the most reasonable ways to estimate the number of classes of specific binding sites from ligand-binding data.

The AIC method was originally developed as a statistical model selection method for an autoregressive model of time series (9). This method has been widely used in the field of engineering and recently has been successfully applied to estimate the number of exponential terms in analyzing time course data of plasma concentrations of a drug (16). In the present study analyzing ligand-binding data, the AIC method selected the same models as the *F*-test did. In addition, both the methods could find the correct models in analyzing artificially-generated data. These results suggest that the AIC method can be applicable to estimate the number of classes of specific binding sites in analyzing ligand-binding data. The AIC method is free from subjectivity which is encountered in setting a significance limit in

---

**Fig. 2.** Graphical representation of nonlinear regression analysis of artificially-generated ligand-binding data. Data were artificially generated as described in the text and analyzed by the nonlinear least-squares procedure. Units of parameters are arbitrary. (A) Data were generated based on a model with two parameters: $K_d=20$, $B_{\text{max}}=1$, a constant coefficient of variance (CV) $=4\%$. The data were analyzed with models with two or three parameters. The following results were obtained: A model with two parameters (---): $K_d=19.3$, $B_{\text{max}}=0.98$, residual sum of squares (SS) $=4.29$, AIC $=20.0$. A model with three parameters: $K_d=17.7$, $B_{\text{max}}=0.90$, NSB $=0.084\%$, SS $=3.70$, AIC $=20.4$. $F_{23}=1.28$ (not significant). Both the AIC method and the *F*-test method selected the model with two parameters. (B) Data were generated based on a model with three parameters: $K_d=10$, $B_{\text{max}}=1$, NSB $=0.3\%$, CV $=4\%$. The data were analyzed with models with two, three or four parameters. The following results were obtained. A model with two parameters: $K_d=12.5$, $B_{\text{max}}=1.23$, SS $=23.8$, AIC $=6.05$. A model with three parameters (---): $K_d=10.3$, $B_{\text{max}}=1.03$, NSB $=0.254\%$, SS $=9.6$, AIC $=1.39$. A model with four parameters: fitting did not converge. $F_{23}=11.83$ (P $<0.001$). Both the AIC method and the *F*-test method selected the model with three parameters. (C) Data were generated based on a model with four parameters: $K_d=10$, $B_{\text{max}}=1$, $K_d=50$, $B_{\text{max}}=1$, CV $=4\%$. The data were analyzed with models with two, three, four or five parameters. The following results were obtained. A model with two parameters: $K_d=16.8$, $B_{\text{max}}=1.83$, SS $=32.3$, AIC $=59.6$. A model with three parameters: $K_d=14.8$, $B_{\text{max}}=1.66$, NSB $=0.042\%$, SS $=19.3$, AIC $=53.4$. A model with four parameters (---): $K_d=7.2$, $B_{\text{max}}=0.72$, $K_d=49.1$, $B_{\text{max}}=1.29$, SS $=5.482$, AIC $=35.2$. A model with five parameters: $K_d=7.15$, $B_{\text{max}}=0.71$, $K_d=47.5$, $B_{\text{max}}=1.30$, NSB $=0.0015\%$, SS $=5.476$, AIC $=37.2$. $F_{23}=8.756$ (P $<0.001$), $F_{34}=30.25$ (P $<0.001$). $F_{45}=0.012$ (not significant), $F_{24}=29.35$ (P $<0.001$). Both the AIC method and the *F*-test method selected the model with four parameters.
the $F$-test. In performing multiple $F$-tests (as in Fig. 1C and Fig. 2C), it is difficult to give explicitly a required significance level. The AIC method is also simple and does not require any statistical table. From these considerations, it is concluded that the AIC method can be a good alternative to the $F$-test method in estimating the number of classes of specific binding sites from experimental ligand-binding data.

It is obvious that the estimated number of classes of specific binding sites does not necessarily have physical meaning. Classes of sites with dissociation constants sufficiently close to each other could hardly be resolved by any method. The estimated number of classes of binding sites is dependent on the structure of the data such as number, range, accuracy and precision of data. The estimate is optimal in the statistical sense only for a given condition of data. It is also clear that besides the site heterogeneity model, other mechanical or molecular models such as the concerted or sequential cooperativity model, receptor polymerization model, ligand-induced coupling of receptor to a non-binding component, flexible ligand binding, etc. can explain phenomenological cooperativity in ligand binding (17).

References

1 Weder, H.G., Schildknecht, J., Lutz, R.A. and Kesselring, P.: Determination of binding parameters from Scatchard plots: theoretical and practical considerations. Eur. J. Biochem. 42, 475–481 (1974)

2 Munson, P.J. and Rodbard, D.: Ligand: a versatile computerized approach for characterization of ligand-binding systems. Anal. Biochem. 107, 220–239 (1980)

3 Klotz, I.M.: Numbers of receptor sites from Scatchard graphs: facts and fantasies. Science 217, 1247–1249 (1982)

4 Burgisser, E.: Model testing in radioligand/receptor interaction by Monte Carlo simulation. J. Recept. Res. 3, 261–281 (1983)

5 Thakur, A.K., Jaffe, M.L. and Rodbard, D.: Graphical analysis of ligand-binding systems: evaluation by Monte Carlo studies. Anal. Biochem. 107, 279–295 (1980)

6 Feldman, H.A.: Mathematical theory of complex ligand-binding systems at equilibrium. Anal. Biochem. 48, 317–338 (1972)

7 De Lean, A., Hancock, A.A. and Lefkowitz, R.J.: Validation and statistical analysis of a computer modeling method for quantitative analysis of radioligand binding data for mixtures of pharmacological receptor subtypes. Mol. Pharmacol. 21, 5–16 (1982)

8 Feldman, H.A.: Statistical limits in Scatchard analysis. J. Biol. Chem. 258, 12865–12867 (1983)

9 Akaike, H.: A new look at the statistical model identification. IEEE Trans. Automat. Contr. AC-19, 716–723 (1974)

10 Akaike, H.: An information criterion (AIC). Math. Sci. 14, 5–9 (1976)

11 Wilson, S.P. and Kirshner, N.: Isolation and characterization of plasma membranes from the adrenal medulla. J. Neurochem. 27, 1289–1298 (1976)

12 Kamikubo, K., Miura, K. and Fujimura, H.: Elimination of GTP biphasic regulation of synaptosomal adenylate cyclase by manganese and solubilization. Japan. J. Pharmacol. 32, 893–902 (1982)

13 Nozaki, M., Kamikubo, K., Niwa, M., Fujimura, H., Imai, E., Suzuki, Y. and Hori, M.: Chemical features for opiate agonist and antagonist. In Endogenous and Exogenous Opiate Agonist and Antagonist, Edited by Way, E.L., p. 9–12, Pergamon Press, New York (1980)

14 Flecher, R.: A modified Marquardt subroutine for nonlinear least squares. Tech. Report AERE R6799, U.K. Atomic Energy Authority Research Establishment, Harwell (1971)

15 Yamaoka, K., Tanigawara, Y., Nakagawa, T. and Uno, T.: A pharmacokinetic analysis program (MULTI) for microcomputer. J. Pharmacobiodyn. 4, 879–885 (1981)

16 Yamaoka, K., Nakagawa, T. and Uno, T.: Application of Akaike’s information criterion (AIC) in the evaluation of linear pharmacokinetic equations. J. Pharmacokinet. Biopharm. 6, 165–175 (1978)

17 Boeynaems, J.M. and Dumont, J.E.: Outlines for Receptor Theory. Elsevier/North-Holland Biomedical Press, Amsterdam (1980)