

LETTER TO THE EDITOR

Saturation Functions as a Nested Set

In a recent paper, Bardsley & McGinlay (1987) use saturation functions for binding of a ligand to a protein with an unknown number of binding sites to explore the power and reliability of the F test when used outside its range of strict validity. In the course of their discussion they make the interesting and novel observation that binding equations do not form a nested set: increasing the number of terms does not give a model that includes the simpler ones as special cases, and does not necessarily result in a minimum sum of squares smaller than or equal to that obtained with an equation with fewer terms. In their Appendix 5 they prove that this is a real result rather than, for example, a computing artifact.

There is an important sense, however, in which saturation functions do constitute a nested set. For convenience, one often treats the fractional saturation y , defined as the number of occupied binding sites divided by the total number of binding sites whether occupied or not, as the dependent variable, and this is what Bardsley & McGinlay (1987) do in their paper. However, y is never measured directly; it must be calculated by dividing the observed degree of binding by the degree of binding at saturation. This limit is commonly obtained in either of two ways which differ in their consequences for model testing. One way is to estimate the limit by extrapolation of the observed behaviour to infinite ligand concentration, and in this case the value of y is model-independent and it is quite proper to treat it as dependent variable in the statistical calculations. (Nonetheless, one might prefer to treat the observed degree of binding as the dependent variable and to include the limit as a parameter to be estimated.) Alternatively, if the purity and molecular mass of the protein are known it is preferable to take as dependent variable the average number of ligand molecules bound per molecule of protein, N_x : this corresponds to ny , where n is the number of binding sites per molecule. In this case y is model-dependent, and when testing the significance of including an extra term in the binding equation, therefore, i.e. increasing n from i to $i+1$, one must not keep all the y values unchanged but must divide them by $(1+1/i)$.

One can avoid this complication by taking N_x explicitly as dependent variable rather than y . With N_x as dependent variable, any binding function may be written as the sum of a series of simple binding terms:

$$N_x = \sum_{i=1}^n \frac{K_i x}{1 + K_i x}. \quad (1)$$

If the n binding sites are independent (non-interacting) the parameters K_i are real numbers and define the association constants for the separate sites. More generally, the K_i are complex numbers and it is more convenient for computational purposes to define the model in terms of the coefficients of the binding polynomial $p_n(x)$,

the product of the denominators of the terms in eqn (1), i.e.

$$p_n(x) = \prod_{i=1}^n (1 + K_i x) = 1 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_n x^n. \quad (2)$$

The coefficients α_i of the binding polynomial are real numbers regardless of whether the K_i are real or complex. However, the properties of the sum of squares as n changes are most conveniently discussed in relation to eqn (1), and it is then unimportant whether the parameters are real or not.

If the binding function is written as in eqn (1), it is obvious that the models for different n values do constitute a nested set, i.e. that setting some of the K_i values to zero will always produce a valid model with a smaller value of n . Moreover, increasing n must always give a model that fits at least as well as all of the models with smaller n .

In conclusion, I should emphasize that I believe that although this observation may require some of the details of the work of Bardsley & McGinlay (1987) to be revised, it is unlikely to affect their main conclusions, for which binding functions provide an illustration rather than the central issue. In the particular case of binding of ligands to proteins, the discrimination power of the F test is probably less than indicated by the results of Bardsley & McGinlay (1987) when y is taken as dependent variable, because of the implicit extra parameter represented by the degree of binding at saturation.

For experiments where N_x is the dependent variable, it is certainly easier to discriminate between models than it is if the value at saturation has to be estimated. The simple fact that some of the observed N_x values exceed 2 or 3 may often be sufficient to exclude these from consideration as possible values for n . Consider, for example, two investigations that were examined in detail in a previous discussion of the computational and statistical aspects of binding data (Cornish-Bowden & Koshland, 1970). In the data of Cook & Koshland (1970) for the binding of NAD^+ to yeast glyceraldehyde 3-phosphate dehydrogenase, four of the observed values of N_x exceeded 3, indicating that n must be at least 4. Similarly, in the data of Rossi-Fanelli *et al.* (1961) for the binding of oxygen to haemoglobin, the observed N_x values extended to about 3.8, again showing that values n less than 4 could be excluded from consideration.

Centre de Biochimie et de Biologie Moléculaire
Centre National de la Recherche Scientifique,
31 chemin Joseph-Aiguier,
B.P. 71, 13402 Marseille, France

ATHEL CORNISH-BOWDEN

(Received 24 July 1987)

REFERENCES

- BARDSLEY, W. G. & MCGINLAY, P. B. (1987). *J. theor. Biol.* **126**, 183.
 COOK, R. A. & KOSHLAND, D. E., JR. (1970). *Biochemistry* **9**, 3337.
 CORNISH-BOWDEN, A. & KOSHLAND, D. E., JR. (1970). *Biochemistry* **9**, 3325.
 ROSSI-FANELLI, A., ANTONINI, E. & CAPUTO, A. (1961). *J. biol. Chem.* **236**, 397.