Biophysical Chemistry, 34 (1989) 19–28
Elsevier

BIOCHE 01368

Information theory and the analysis of ligand-binding data

Enrico Di Cera [a], Francesco Andreasi Bassi [a] and Stanley J. Gill [b]

[a] *Istituto di Fisica, Universita' Cattolica, Largo F. Vito I, 00168 Roma, Italy and* [b] *Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215, U.S.A.*

Received 15 February 1989
Revised manuscript received 11 April 1989
Accepted 25 April 1989

Data analysis; Thermodynamics; Hemoglobin

The phenomenological principles of information theory are used in the analysis of ligand-binding phenomena in biological macromolecules. Information maps are constructed to visualize regions of ligand chemical potential with maximum amount of information and to devise suitable experimental strategies therefrom. Extensive simulation studies and analysis of experimental data also point out the properties of information used as a weighting procedure in nonlinear least-squares analyses.

1. Introduction

The analysis of ligand-binding phenomena in biological macromolecules is a critical step in understanding the functional properties of biological systems. In the case of systems at equilibrium the analysis can be cast in terms of a partition function and straightforward application of statistical thermodynamics yields all the relevant relations for the quantitative description of the system [1]. The thermodynamic approach is perfectly general, based as it is on measurements of ligand chemical potentials and free-energy changes associated with binding reactions. It guarantees a most rigorous analysis of ligand-binding phenomena in phenomenological terms.

A key role in statistical thermodynamics is played by information theory [2,3]. This approach too is phenomenological and draws attention to the problem of how information is stored in a system and exchanged with the surroundings. Yet, it does not seem to have received a great deal of attention in the analysis of ligand-binding phenomena in biological macromolecules. In this paper, we wish to outline some relevant features of information theory, the way they apply to ligand-binding phenomena and their analysis. The approach taken here draws attention to the critical evaluation of experimental strategies usually followed in the study of binding phenomena. The analysis cast in terms of phenomenological principles has the advantage of being general enough to allow consideration of linkage effects under a much broader perspective.

2. The information approach

Consider a hypothetical system composed of a biological macromolecule, M, which can exist in several different ligation states with respect to a ligand X. At equilibrium each state of the macromolecule enters the definition of a partition function, Z, as follows

$$Z = \sum_{j=0}^{t} \exp(-E_j/RT) \qquad (1)$$

Correspondence address: E. Di Cera, Istituto di Fisica, Universita' Cattolica, Largo F. Vito 1, 00168 Roma, Italy.

where R and T have the usual meaning, t represents the number of binding sites for ligand X, while E_j is the energy level associated with the j-th configuration and is equal to

$$E_j = \Delta G_j - j\mu \tag{2}$$

where ΔG_j is the (standard) free energy change for the reaction $M + jX = MX_j$ and μ is the ligand chemical potential.

From the partition function one readily calculates the probability p_j for the macromolecule being in state j as $p_j = \exp(-E_j/RT)/Z$. We now define the work V_j to take the system as a whole from a given distribution of ligation states into state j. This work is given by the difference between the basic, or standard, free energy of state j and its chemical potential [4,5], is always positive, and is equal to the free-energy change for the transition $p_j \to 1$, i.e.,

$$V_j = -RT \ln p_j \tag{3}$$

The quantity V_j is also a gross free energy [4], since it depends on the properties of the ensemble of ligation states. In this respect it reflects a property of the system as a whole.

Since there are $t + 1$ possible ligation states, so one has as many terms such as eq. 3. If each V_j is weighted by the corresponding probability p_j, one obtains the average energy $\langle V \rangle$ needed for the system to exist in only one of its $t + 1$ possible states, i.e.,

$$\langle V \rangle = \sum_{j=0}^{t} p_j V_j = -RT \sum_{j=0}^{t} p_j \ln p_j \tag{4}$$

The physical interpretation of eq. 4 is straightforward. Since when $p_j \to 1$ one has $V_j \to 0$ and $\langle V \rangle \to V_j$, then $\langle V \rangle$ is the minimum work needed to collapse the actual distribution of the ensemble of states into a single state. The energy $\langle V \rangle$, when expressed in RT units, is equal to the information I of the system, i.e.,

$$I = -\sum_{j=0}^{t} p_j \ln p_j \tag{5}$$

This derivation of the function I does not differ conceptually from the 'classical' one [2,3], although it emphasizes the free-energy changes associated with the underlying binding reactions taking place in the system.

Consider now the prototypic case of two ligands, X and Y, that bind to the macromolecule and interfere with each other. Such a system represents the simplest example of interest in a thermodynamic description of linkage effects [6]. If there are t sites for ligand X and s sites for ligand Y, then the partition function contains $(t + 1)(s + 1)$ terms and is equal to

$$Z = \sum_{i=0}^{t} \sum_{j=0}^{s} \exp\left[\left(i\mu_X + j\mu_Y - \Delta G_{ij}\right)/RT\right] \tag{6}$$

where μ_X and μ_Y denote the chemical potentials of the two ligands, and ΔG_{ij} is the (standard) free-energy change for the reaction $M + iX + jY = MX_iY_j$. The probability p_{ij} related to state ij is given by $\exp[(i\mu_X + j\mu_Y - \Delta G_{ij})/RT]/Z$, and the information I_{XY} in the presence of both ligands is

$$I_{XY} = -\sum_{i=0}^{t} \sum_{j=0}^{s} p_{ij} \ln p_{ij} \tag{7}$$

As in the case of a single ligand, the information I gives the minimum energy, in RT units, necessary to take the whole system into one of its $(t + 1)(s + 1)$ possible ligation states. The general form of the function I for an arbitrary number of ligands is easily derived from eq. 7.

The second ligand increases the amount of information stored in the system by expanding the ensemble of ligation states from $t + 1$ to $(t + 1)(s + 1)$. The maximum increase is observed when the two ligands are independent due to additivity of the information stored in the system. Consequently, linkage can only reduce the amount of information in the ensemble of ligation states. The demonstration is as follows. We calculate the minimum energy needed to take the system in one of the $t + 1$ states defined with respect to ligand X as

$$I_X = -\sum_{i=0}^{t} \sum_{j=0}^{s} p_{ij} \ln \sum_{j=0}^{s} p_{ij} \tag{8}$$

and similarly I_Y is given by

$$I_X = -\sum_{i=0}^{t}\sum_{j=0}^{s} p_{ij} \ln \sum_{i=0}^{t} p_{ij} \tag{9}$$

Combination of eqs. 7–9 yields the well-known inequality [2]

$$I_{XY} \leq I_X + I_Y \tag{10}$$

along with

$$I_{XY} \geq I_X \tag{11}$$

These relations show that linkage, either positive or negative, always decreases information and the absence of linkage leaves information unaltered. Generalization to an arbitrary number of ligands is straightforward. Linkage gives rise to two effects that oppose each other. It increases the number of possible ligation states, and hence information (see eq. 11), due to the presence of a second ligand. On the other hand, this increase does not yield the maximum amount of information that the system is able to store when the two ligands are independent (see eq. 10). This is because linkage necessarily selects only particular distributions among all those possible that are contained in the ensemble.

Linkage is also a measure of cooperativity, i.e., the interactions among the sites of a macromolecule when only one type of ligand is bound. Cooperativity always affects information as it 'biases' the distribution of the states of the macromolecule toward the unligated and fully ligated forms. The reduction of the number of states significantly populated is translated into a decrease in information. A thermodynamic measure of cooperativity is the binding capacity which gives the change in the number of moles of a given ligand bound to the macromolecule due to a change in its chemical potential. The principle of the binding capacity provides the magnitude and direction of the change in cooperativity of a given ligand due to a second linked ligand. This principle states that the binding capacity, and hence cooperativity, is a maximum when the linked ligand is buffered [7]. Consequently, unbuffering the system results in increased information.

When ligand binding is linked to change in the aggregation state of the macromolecule the ensemble of states is also expanded along two coordinates. The partition function for the general case can be cast in a variety of forms [8,9]. A convenient one is the phenomenological expansion

$$Z = \sum_{i=1}^{s} c_i m^i Z_i \tag{12}$$

where m denotes the concentration of free monomer, which is assumed to contain a single binding site, s is the maximum aggregation state, the c_i are equilibrium constants of aggregation for the unligated forms, and the Z_i are the partition functions for each aggregation state. They are defined as

$$Z_i = \sum_{j=0}^{i} \exp\left[\left(j\mu - \Delta G_{ij}\right)/RT\right] \tag{13}$$

where ΔG_{ij} is the (standard) free-energy change for the reaction $M_i + jX = M_iX_j$. The probability of each state ij is calculated as

$$p_{ij} = c_i m^i \exp\left[\left(j\mu - \Delta G_{ij}\right)/RT\right]/Z \tag{14}$$

and the concentration of free monomers is obtained from the condition of mass conservation as a root of the s-degree polynomial

$$\partial Z/\partial \ln m = \sum_{i=1}^{s} i c_i m^i Z_i = m_t \tag{15}$$

with m_t equal to the total concentration of monomers. The information content of such a system can be defined along two coordinates, i.e., the ligand activity x and m_t, referring to the ligation and aggregation states, respectively.

3. Information maps

The principles derived above regarding the information content of a system composed of a macromolecule and its ligands can be used in the analysis of ligand-binding data. The ensemble of states which defines the partition function contains all the species that make a contribution to the experimental measurements. Determination of

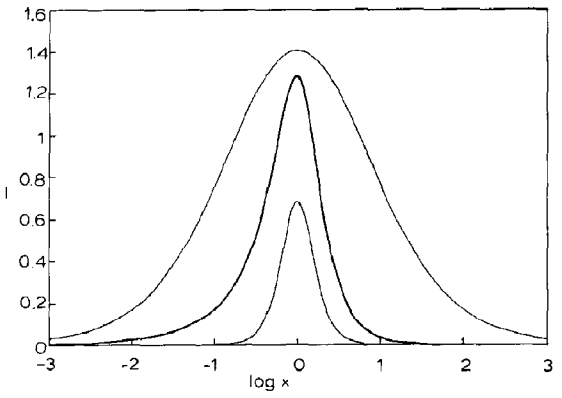

Fig. 1. Information of O_2 binding to human hemoglobin A_0 as a function of O_2 fractional saturation (middle curve) under experimental conditions of 12 mM heme, 0.1 M phosphate, pH 7.4, and 25°C [10]. Information of a noncooperative (upper curve) and fully cooperative (lower curve) tetramer is also shown for comparison.

the particular distribution of ligation intermediates under given experimental conditions is critical in determining the functional properties of the macromolecule. Since the distribution is also related to the information stored in the system, one should pay much attention to where the maximum information is located and how it is distributed in the binding study being considered. This leads to consideration of information maps.

In fig. 1 we show the function I for the case of O_2 binding to human hemoglobin A_0 [10]. The information content of a noncooperative and fully cooperative tetramer is also shown for comparison. When the sites are all alike and independent (upper curve) the amount of information stored in the system is a maximum. In the positively cooperative case of O_2 binding to human hemoglobin, the amount of information (middle curve) is drastically reduced and approaches that of a fully cooperative tetramer (lower curve). This represents an intrinsic and basic difficulty in the resolvability of the free-energy changes due to O_2 binding from analysis of experimental data, as also pointed out previously from a different standpoint [11]. In all cases, however, the function I reaches its maximum in the middle region of the saturation curve. It is here that all ligation intermediates make their appearance and significantly contribute to the information stored in the system. One also observes that little, if any, information resides in the end regions of the saturation range where the macromolecule predominantly exists in either the unligated or fully ligated form. The middle region of the binding curve should be scrutinized carefully in the study of a cooperative system such as human hemoglobin, and the information map shown in fig. 1 should be best used to establish where experimental measurements need to be concentrated.

When two ligands are present the energetics of ligand binding to the macromolecule is cast in terms of an expanded ensemble of ligation states.

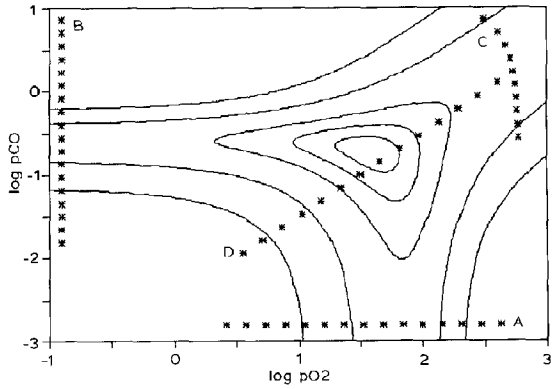

Fig. 2. Information map for the identical linkage of O_2 and CO binding to human hemoglobin A_0 under experimental conditions of 600 μM heme, 0.1 M Hepes, 0.1 M NaCl, 10 mM inositol hexaphosphate, 1 mM Na_2EDTA, pH 6.94, and 25°C [12]. Information levels are shown as continuous lines and correspond to 25, 45, 70, 85 and 95% (innermost curve) of the maximum information. The map shows the differences in information observed upon binding of O_2 and CO separately. The levels occupy a narrower range of partial pressures in the case of CO indicating a higher cooperativity. The maximum amount of information is stored in the system at intermediate values of O_2 and CO partial pressures (inner curves). The asymmetry in the information levels progressively decreases by increasing the partial pressure of both ligands (top right-hand region of the map). Symmetry and asymmetry in this information map respectively match validity and failure of Haldane's laws [13]. The experimental strategy of approaching the study of the CO-O_2 identical linkage is depicted in the map by cuts corresponding to different experiments. (A) O_2-binding curve and (B) CO-binding curve: the cuts are reported in the map for finite values of the CO (A) or O_2 (B) partial pressure solely to allow visualization. (C) Competition experiment. (D) Diagonal experiment.

This also expands the dimensions of information maps where the coordinates are related to different ligands. The information map for the case of O_2 and CO binding to human hemoglobin A_0 [12] is shown in fig. 2 as a contour plot. The information surface is conveniently projected in the plane of ligand chemical potentials, and the levels of interest are shown as continuous lines. The map readily reveals the range of O_2 and CO chemical potentials where information is maximum. A certain symmetry can be seen in the range of high O_2 and CO partial pressures, consistent with a constant value of the partition coefficient and hence validity of the Haldane laws [13] in this region. The symmetry progressively breaks down at low partial pressures due to a failure of the Haldane laws [12]. The map can also be used to check the reliability of the experimental strategies used in dissecting the energetic features of the system. The linkage scheme can be studied in several ways. Information regarding the ligation intermediates with O_2 and CO separately is obtained by studying the binding curve of one ligand in the absence of the other. These two experiments yield two cuts in the map parallel to the two axes in the region where the partial pressure of the other ligand is negligible. These cuts do not cross the region of maximum information, in agreement with the inequality, eq. 11. In the competition experiment, hemoglobin is first saturated with CO and then O_2 is added to displace CO from the binding sites [12]. The cut in the information graph given by this experiment is in a region of higher information content with respect to the previous experiments. The region of maximum information is crossed in the diagonal experiment. In this case, hemoglobin is equilibrated with a mixture of CO and O_2 which is stepwise diluted to keep the difference in the chemical potentials of the two ligands constant. The cut characterizing the experiment shows up in the map as a straight line. One sees that the four experiments altogether provide an efficient strategy for the exploration of the information map in its entirety. The diagonal experiment is critical in dissecting the peculiar energetic features of the ligation intermediates. This fact was first recognized in the experimental approach to the study of the CO-O_2 linkage to human hemoglobin [12], and is now given a sound thermodynamic basis in terms of information theory. A proper design of several diagonal experiments would explore the region of maximum in-

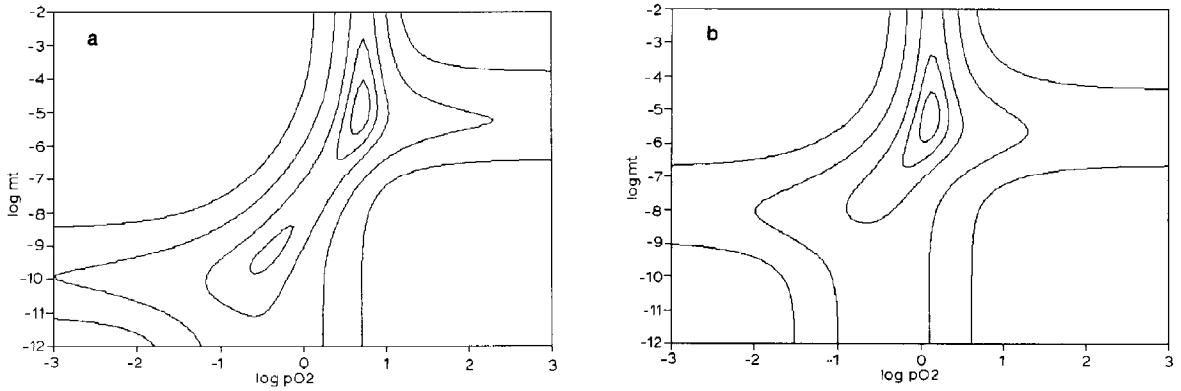

Fig. 3. (a,b) Information map for the O_2-linked dissociation of human hemoglobin A_0 into dimers under experimental conditions of 0.1 M Tris, 0.1 M NaCl, 1 mM Na_2EDTA, pH 7.4, and 21.5 °C (a); or 0.1 M glycine, 0.1 M NaCl, 1 mM Na_2EDTA, pH 9.5, and 21.5 °C (b) [14]. The two coordinates in the map are the O_2 partial pressure and the molar heme concentration. Information levels correspond to 25, 45, 70, 85 and 95% (innermost curve) of the maximum information. The map shows the limiting regions where only dimers or tetramers exist and their differences due to cooperativity. It also points out the existence of a high information peak in the nanomolar concentration range (a), which is inaccessible experimentally. The peak disappears at pH 9.5 (b). In the millimolar-micromolar range, which is the region explored experimentally [14], it is interesting to note that information does not drop to zero at high O_2 partial pressures due to the coexistence of fully ligated dimers and tetramers. A similar effect is observed at low O_2 partial pressures in the nanomolar-picomolar range.

formation in great detail and possibly improve resolution of the free-energy changes associated with the ligation intermediates.

Another demonstration of the possible relevance of information theory arising in connection with ligand binding thermodynamics is given by the graph shown in fig. 3. The map depicts the information content of the O_2-linked dissociation of human hemoglobin tetramers into dimers at two different pH values [14]. Here, the total heme concentration, m_t, gives the second coordinate of the system and eq. 15 has an exact solution [15], which yields the concentration of free heme, m, and hence the probability of each state. The map is somewhat more complex than that shown in fig. 2. The two extreme cases are observed at low and high concentrations, where hemoglobin is predominantly in its dimeric or tetrameric state. At intermediate concentrations the dimer-tetramer equilibria result in the coexistence of unligated or fully ligated states due to both aggregation forms. The information increases accordingly and, contrary to what is observed at the limiting concentration regions, it does not drop to zero at very low and high values of O_2 chemical potential. This is an important difference with respect to ligand-binding studies performed on purely tetrameric hemoglobin [10,12]. The experimental approach to the study of the dimer-tetramer linkage scheme is focussed on the millimolar-micromolar concentration range. This range is effective in exploring the region of maximum information content. There is, however, another region of high information content in the nanomolar range (see fig. 3a), but it is inaccessible experimentally. Interestingly, this region disappears at pH 9.5 where the alkaline Bohr effect is over (see fig. 3b).

4. Information and the analysis of ligand-binding data

The principles of information theory also contain some possible guidelines for the analysis of experimental data. In this regard we have examined the properties of information used as a weighting function in nonlinear least-squares analysis. Such a function assigns to each experimental point a differential weight based on its intrinsic information content. Although the use of information weighting seems to be justified by the considerations outlined in the preceding sections, it is nevertheless at variance with a well-known statistical principle which states that experimental points should be weighted according to the intrinsic error measured experimentally. When the errors are uniformly distributed, as usually observed experimentally in ligand-binding studies, all points should receive uniform weight. In order to clarify this point, we have studied the effect of information weighting on the resolvability of free-energy changes associated with ligand binding by Monte Carlo simulations.

A large set of binding curves were simulated in the form of optical absorbance values using the expression [16]

$$A(x_i) = A_0 + A_T\theta(x_i) \tag{16}$$

where $A(x_i)$ is the i-th absorbance value corresponding to ligand activity x_i and A_0 denotes the value of A in the absence of ligand, A_T being the total absorbance change observed when the fractional saturation, θ, is equal to 1. For a nondissociating macromolecule the function θ is equal to

$$\theta = RT(\mathrm{d}\ln Z/\mathrm{d}\mu)/t = (\mathrm{d}\ln Z/\mathrm{d}\ln x)/t$$
$$= \sum_{i=0}^{t} i\beta_i x^i / \sum_{i=0}^{t} t\beta_i x^i \tag{17}$$

where x is the ligand activity, and β_i, the overall equilibrium constant for the reaction $M + iX = MX_i$. Two simple cases were simulated: binding to a monomeric macromolecule for which

$$\theta = \beta_1 x/(1 + \beta_1 x) \tag{18}$$

and to a dimeric macromolecule expressing maximum cooperativity for which

$$\theta = \beta_2 x^2/(1 + \beta_2 x^2) \tag{19}$$

500 binding curves of 40 data points each were generated by adding a pseudorandom error, δ, to eq. 16 on the order of 0.003, as typically found in high-precision binding studies [10,12,14], to yield

$$S(x_i) = A_0 + A_T\theta(x_i) + \delta_i \tag{20}$$

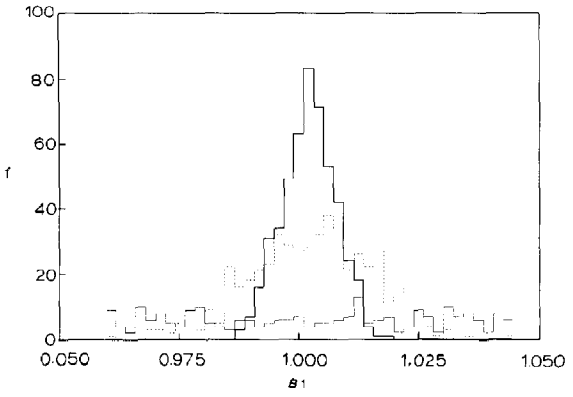

Fig. 4. Distributions of the β_1 values obtained by Monte Carlo simulations in the case of a monomeric macromolecule. The distributions were constructed by collecting the best-fit values of 500 binding curves simulated according to eq. 20 as described in the text, with $\beta_1 = 1$, $A_0 = 1$ and $A_T = 1$. The distributions obtained with uniform ($w_i = 1$ for all i) and information weighting (eq. 22) are shown by bold and dashed lines, respectively, and are consistent with the simulated values. The distribution obtained by weighting the data according to the inverse of information (eq. 23) is shown by a thin line. It is severely asymmetric and the value of β_1 is poorly defined.

The data were then fitted by minimizing the function

$$\Phi = \sum_{i=1}^{n} w_i \left[S(x_i) - F(x_i) \right]^2 \qquad (21)$$

where $n = 40$, and F is the fitting function equal to eq. 16, with θ given by eq. 18 or 19. Three different expressions were used for the weights w_i in eq. 21. First, a uniform weighting procedure, such that $w_i = 1$ for all i, was used and the best-fit values of the equilibrium constants and the optical parameters A_0 and A_T were collected for statistical analysis. In the case of a monomeric macromolecule, the distribution of β_1 values was found to be Gaussian, with a mean identical to the simulated value, and a standard deviation consistent with the magnitude of the simulated error (see fig. 4). The data were then fitted by using 'information weighting', so that the points in the middle region of the binding curve received more weight than the end points. The weighting function was in this case

$$w_i = \left\{ \theta(x_i) \left[1 - \theta(x_i) \right] \right\}^2 \qquad (22)$$

which closely approximates eq. 5. The distribution found for the β_1 values is in good agreement with that obtained using a uniform weighting, and is shown in fig. 4. The interesting conclusion is that the differential information weighting given by eq. 22 does not bias the statistics and provides results consistent with uniform weighting. Similar results were obtained in the case of the dimeric macromolecule with maximum cooperativity (see fig. 5). To explore further the effect of information on the

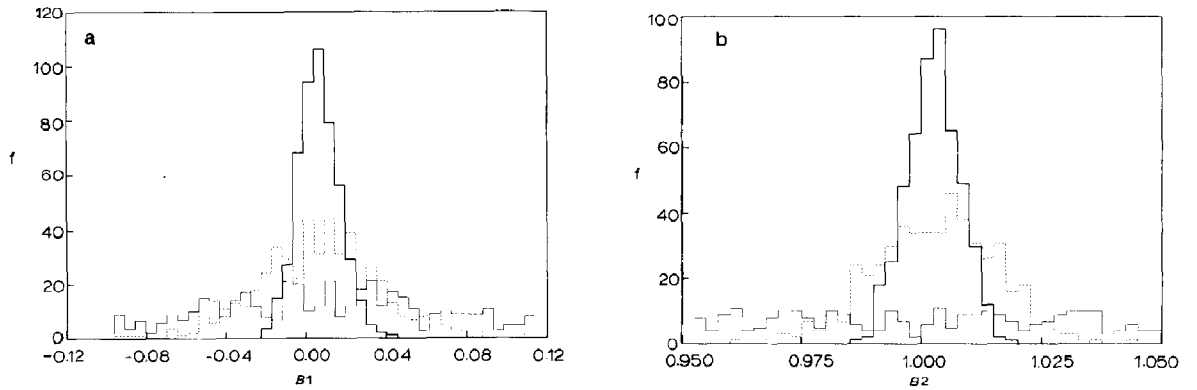

Fig. 5. (a,b) Distributions of the β_1 (a) and β_2 (b) values obtained by Monte Carlo simulations in the case of a fully cooperative, dimeric macromolecule. The distributions were constructed by collecting the best-fit values of 500 binding curves simulated according to eq. 20 as described in the text, with $\beta_1 = 0$, $\beta_2 = 1$, $A_0 = 1$ and $A_T = 1$. As in the case of fig. 4, the distributions obtained with uniform (bold line) and information weighting (dashed line) are consistent with the simulated values, while weighting according to the inverse of information (thin line) yields biased and poorly defined parameter estimates.

resolvability of the equilibrium constant for the simulated cases described above, the data were also fitted by using a third weighting function as follows

$$w_i = \{\theta(x_i)[1-\theta(x_i)]\}^{-2} \qquad (23)$$

so that points were weighted according to the inverse of their information. In this case the bias in the distribution of the equilibrium constants is significant, and these parameters are poorly resolved as indicated by the drastic increase in the standard deviation of the distributions (see figs. 4 and 5).

Analysis of the data was also performed by fixing the two optical parameters A_0 and A_T in eq. 16 to the simulated values. Fixing the optical parameters was found to increase the correlation between information and uniform weighting (see fig. 6), and did not yield biased parameter estimates. The same result was obtained in the case of the dimeric macromolecule. This indicates that when the values of the optical parameters are accurately determined, a few points in the middle region of the binding curve would guarantee the resolvability of equilibrium constants.

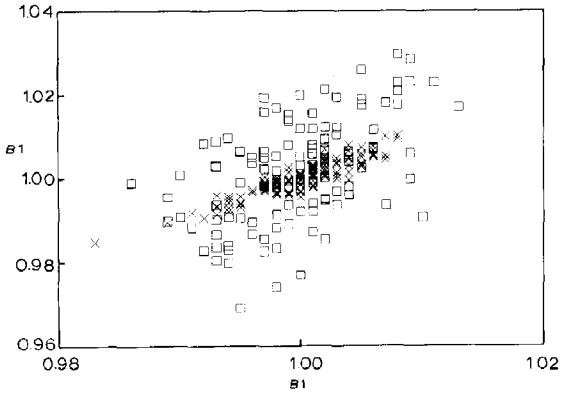

Fig. 6. Correlation plot between the values of β_1 obtained with uniform (abscissa) and information weighting (ordinate) for the case of a monomeric macromolecule. The plot was derived by floating (□) or fixing (×) the optical parameters A_0 and A_T in the fitting procedure to their simulated values ($A_0 = 1$; $A_T = 1$). Only 120 out of 500 points are shown for the sake of clarity. The same result is obtained for a fully cooperative, dimeric macromolecule. The value of the correlation coefficient is 0.49 (□) and 0.92 (×). The correlation between uniform weighting and weighting according to the inverse of information is less than 0.15, either floating or fixing the optical parameters.

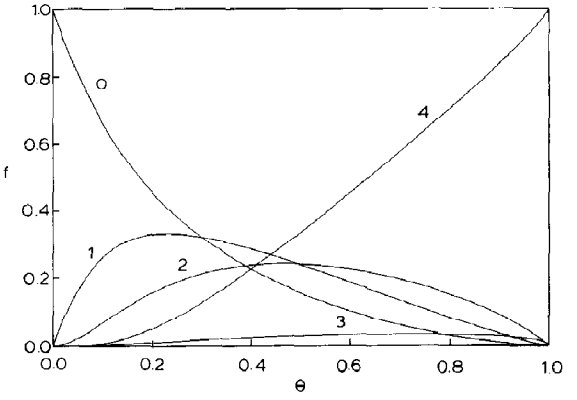

Fig. 7. Species fraction plot for the five intermediates in the O_2 binding reaction to human hemoglobin A_0 obtained by weighting the data according to the inverse of information (eq. 23). The triply ligated species are populated up to 4% due to the incorrect weighting procedure used. Information weighting (eq. 22), on the other hand, yields results consistent with those obtained by uniform weighting ($w_i = 1$ for all i) [16], and confirms a negligible population of triply ligated species.

The three different weighting procedures were then tested in the analysis of real experimental data obtained for the O_2-binding reaction of human hemoglobin. Information weighting yields results consistent with those obtained with uniform weighting and, in particular, it confirms a negligible value for the triply ligated species. Weighting according to the inverse of information interestingly results in a substantial overestimation of the triply ligated species, as shown in fig. 7 by the species fraction plot. Such an effect is similar to that found in the simulation study where the value of β_1 for the dimer was positively biased when the data were weighted against information.

5. Discussion

The analysis and interpretation of ligand-binding phenomena can be considerably broadened by taking into account the amount of information stored in the system. This quantity can be seen as

the minimum energy required to drive the system into one of its possible states, and can be correlated to the partition function and the ligation intermediates. The information approach draws attention to the relevance of studying ligand-binding phenomena in terms of linkage effects. Linkage increases the information due to a single component by expanding the number of states due to the effect of a second ligand. However, it should be borne in mind that the overall information of the system is always increased and never maximized by linkage effects, independent of their nature.

The conclusions drawn here apply equally well to the case of linkage between chemical and physical phenomena, where the dominant effects are measured as heat capacity or volume changes of the system. In any case, the analysis of the system cast in terms of the partition function yields relations such as eq. 5, and the states of the macromolecule refer to ligation intermediates expanded along the functional or denatured forms that may be detected in a thermal transition profile. As in the case of chemical linkage, cooperativity and linkage result in a reduced amount of information stored in the system.The ensemble of states is expanded to take into account both the ligation and thermally distinguishable states and can conveniently be encapsulated by information maps. Application of these concepts can be extended to ligand-binding phenomena at the level of individual sites, where the existence of local potentials leads to linkage effects isomorphic with those involving different ligands [17], or else to the description of radiation action on living cells [18].

The use of information in nonlinear least-squares analysis also deserves consideration, since it shows the operationally useful aspects of the theory. As we have seen, information weighting always yields results consistent with uniform weighting, as assessed by either Monte Carlo calculations and analysis of real experimental data. This fact has an obvious practical application, as it suggests that attention should be concentrated on the region where information is a maximum and that experimental measurements should be made predominantly, if not only, in these regions. On the other hand, weighting according to the inverse of information always results in biased and poorly resolved parameter estimates. This aspect should be given much consideration when dealing with data analysis and can be of critical importance in the study of highly cooperative systems, such as human hemoglobin, where the population of intermediates is low [19]. It should be mentioned that the weighting procedure given in eq. 23 has long been used in the analysis of oxygen binding to human hemoglobin [20], and has yielded, not surprisingly, incorrect estimates of the intermediates.

Information theory thus provides phenomenological principles to correctly approaching the analysis of ligand-binding phenomena, either in thermodynamic or in operational terms. As we have seen, construction of information maps may effectively probe the experimental strategy followed in the study of ligand-binding phenomena. The maps can be used to interpret and analyse the dominant linkage effects in a much broader perspective. They have the advantage of bringing together the basic features of the theoretical model under consideration and the experimental measurements, so that the relevant properties of the system are encapsulated in a more comprehensive framework. The operationally useful aspects of information theory can be brought out in this connection. The approach described in this study can be used to assess the range of ligand activity which effectively needs to be explored by experimental measurements. The advantage of such an approach is obvious, especially when the functional properties of a biological macromolecule cannot be tested over a wide range of ligand activity.

Acknowledgements

We are grateful to Drs. Gary K. Ackers, Giuseppe Arcovito, Patrick R. Connelly, Michael L. Johnson and Jeffries Wyman for helpful suggestions and critical comments. This work was supported by NIH grant HL22325 (S.J.G.), and by MPI and CNR (E.D.C. and F.A.B.).

References

1 T.L. Hill, Statistical thermodynamics (Addison-Wesley, Reading, MA, 1960).
2 C.E. Shannon and W. Weaver, The mathematical theory of communication (University of Illinois Press, Urbana, IL, 1949).
3 L. Brillouin, Science and information theory (Academic Press, New York, 1962).
4 T.L. Hill, Free energy transduction in biology (Academic Press, New York, 1977).
5 T.L. Hill and E. Eisenberg, Q. Rev. Biophys. 14 (1981) 463.
6 J. Wyman, Adv. Protein Chem. 19 (1964) 233.
7 E. Di Cera, S.J. Gill and J. Wyman, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 449.
8 J. Hermans and S. Premilat, J. Phys. Chem. 79 (1975) 1169.
9 G. Schwarz and H. Klose, Eur. J. Biochem. 29 (1972) 249.
10 S.J. Gill, E. Di Cera, M.L. Doyle, G.A. Bishop and C.H. Robert, Biochemistry 26 (1987) 3995.
11 S.J. Gill, P.R. Connelly, E. Di Cera and C.H. Robert, Biophys. Chem. 30 (1988) 133.
12 E. Di Cera, M.L. Doyle, P.R. Connelly and S.J. Gill, Biochemistry 26 (1987) 6494.
13 J.S. Haldane and J.L. Smith, J. Physiol. 22 (1897) 231.
14 A.H. Chu, B.W. Turner and G.K. Ackers, Biochemistry 23 (1984) 604.
15 G.K. Ackers and H.R. Halvorson, Proc. Natl. Acad. Sci. USA 71 (1974) 4312.
16 E. Di Cera and S.J. Gill, Biophys. Chem. 29 (1988) 351.
17 E. Di Cera, J. Theor. Biol. 136 (1989) 167.
18 E. Di Cera, F. Andreasi Bassi and G. Arcovito, Biophys. Chem. 33 (1989) 91.
19 S.J. Gill, E. Di Cera, M.L. Doyle and C.H. Robert, Trends Biochem. Sci. 13 (1988) 465.
20 K. Imai, Allosteric effects in hemoglobin (Cambridge University Press, Cambridge, 1982).