Biophysical Chemistry, 45 (1992) 171–179
Elsevier Science Publishers B.V., Amsterdam

BIOCHE 01709

Unique and independent parameters (UIP) formulation for thermodynamic models of complex protein–ligand systems

William G. Gutheil and Charles E. McKenna

Department of Chemistry, University of Southern California, University Park, Los Angeles, CA 90089-0744 (USA)

(Received 18 September 1991; accepted in revised form 22 July 1992)

Abstract

A method for reformulating the thermodynamic (ΔG) description of complex equilibria is presented. The purpose of this reformulation is to take a system of N complexes which is completely defined by N ΔGs, and recast it in terms of a new set of N ΔGs. This reformulation is an extension of the concept of interaction energy (J. Wyman, Adv. Protein Chem. 19 (1964) 223–286). The new ΔGs obtained by this reformulation reflect the intrinsic properties of the binding sites and the hierarchical nature of potential interactions between them. A simple set of rules are developed which allow for the description of complex protein–ligand binding schemes and these rules are used to derive schemes for hemoglobin O_2 binding. This reformulation represents the foundation for the theoretical description of the coupling of energy in protein–ligand systems as illustrated by the theoretical analysis of allosterism in a dimeric protein presented in the following paper. This reformulation also provides the foundation for the analysis of data pertaining to complex equilibria.

Keywords: Cooperativity; Protein–ligand binding; Thermodynamics

1. Introduction

The understanding of protein–ligand binding serves as the foundation for our understanding of nearly all protein mediated phenomena. The ability of proteins to couple their interactions with ligands (cooperatively) is well recognized in biochemistry and serves to optimize the function of proteins (for example the oxygen transporting ability of hemoglobin (Hb)), to allow for the regulation of metabolic pathways (for example the regulation of pyrimidine biosynthesis by aspartate transcarbamoylase), and for the regulation of gene expression (for example the regulation of expression of tryptophan synthesizing enzymes by the tryptophan repressor) [1]. Enzymatic catalysis can in fact be considered as the coupling of ligand binding energy with the lowering of the transition state energy of a reaction [2]. The manifestations of such phenomena are widespread and are generally referred to as allosteric or cooperative interactions. There is an extensive literature on all aspects of this subject reflecting its importance to an understanding of many biochemical phenomena.

Correspondence to: Wm.G. Gutheil, Department of Biochemistry, Tufts University School of Medicine, 136 Harrison Ave., Boston MA 02111 (USA). Tel.: (617) 956-6882, Fax: (617) 956-6409.

A serious impediment to progress in understanding cooperativity is the limitation of current methods of formulating complex physical models for either theoretical analysis or for the analysis of data. For example, the two major physical models for allosteric interaction, the Monod–Wyman–Changeux (MWC) [3] and the Koshland–Nemethy–Filmer (KNF) [4] models, are both subsets of a more complete thermodynamic model which includes all possible conformational and ligand bound states [5]. It would be of considerable theoretical interest to devise a method which would unify both of these extreme views into a common model. For the rigorous analysis of equilibrium binding data pertaining to protein ligand systems the most sophisticated method available is non-linear least squares fitting of data to the Adair equation [6–8]. This method of data analysis is hampered by the number of terms to be fit for complicated systems. There is also continuing debate concerning the statistical and physical significance of the resulting parameters [7,8].

This paper presents a method for describing a complex protein–ligand binding scheme in terms of a set of unique and independent ΔGs, suitable as parameters for both the theoretical analysis of complex thermodynamic models and for the statistical analysis of data pertaining to such models, and discusses what physical interpretation can be ascribed to these parameters. This reformulation is essentially an extension of the concept of interaction energy [9,10]. The presentation of these ideas will begin from first principles to facilitate an understanding of this reformulation. The application of this reformulation to the rigorous analysis of cooperativity in a dimeric protein is presented in the following paper [11].

2. Theory and discussion

2.1 Simple systems and statistical terms relating macroscopic and microscopic constants

The simplest protein–ligand system is a one site protein

$$P + L \xrightarrow{K_a} PL \tag{1}$$

where K_a is the association constant and is defined as

$$K_a = \frac{[PL]}{[P][L]} \tag{2}$$

This discussion is based on equilibrium thermodynamics with the assumption of ideal behavior. In terms of the Gibbs free energy of association

$$\Delta G_a = -RT \ln K_a \tag{3}$$

The ΔGs discussed throughout this paper can be treated as any other ΔGs and we will not discuss further expansion of this parameter, for example into ΔHs, ΔSs, and ΔC_ps for temperature dependent studies, except to note that the reformulation presented here can be easily expanded to handle such cases [12].

There are several alternative formulations available to describe equilibrium systems. In the following discussion association parameters (K_as and ΔG_as) will be used. The corresponding dissociation constants could also be used. The relationships between these are as follows

$$K_d = 1/K_a, \quad \Delta G_d = -\Delta G_a,$$

$$\Delta G_a = -RT \ln K_a = RT \ln K_d \tag{4}$$

A distinction must be drawn between intrinsic or microscopic constants and extrinsic or macroscopic constants. The former describe the behavior of individual sites on the protein while the later describe the aggregate behavior of the system in the experimental frame of reference. To illustrate this point consider the system diagrammed in Fig. 1. In this model there are two identical binding sites for ligand L on protein P. The Ks and ΔGs shown in Fig. 1 are designated as either intrinsic association constants or macroscopic association constants. The intrinsic constants describe the behavior of each site independently and are associated with the fundamental properties of the individual sites. However, on a macroscopic level we cannot distinguish between PL and LP. What we observe macroscopically is PL + LP which is described by the macroscopic association constants shown in Fig. 1C and 1D.

The relationships between the macroscopic and intrinsic association constants in this case are

$$K_1^m = 2K_1^i \quad \text{and} \quad K_2^m = \tfrac{1}{2}K_2^i \tag{5}$$

The simple rationale for such a result is that there are two ways for the first ligand to associate on and only one way to dissociate off so the observed association constant will be twice the intrinsic association constant. A similar argument applies to K_2^m. This point is easily derived algebraically and is important in maintaining the internal consistency of the mathematical formulation of such systems [13]. The equivalent ΔG formulation of the above result is also included in Fig. 1. The $RT\ \ln(x)$ terms are equivalent to the relationships between the intrinsic and macroscopic equilibrium constants discussed above in logarithmic form and can be considered as the entropic contribution of multiple binding sites.

2.2 Cooperativity and interaction terms

Cooperativity and allosteric interactions occupy an important place in a wide variety of

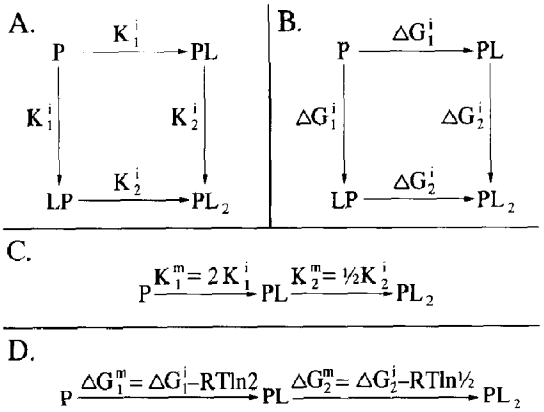


Fig. 1. Equivalent thermodynamic formulations for a homotropic two site protein. Panel A: microscopic formulation in terms of intrinsic equilibrium constants. Panel B: formulation in terms of intrinsic ΔGs. Panel C: equilibrium constant formulation defining the relationship between the intrinsic and macroscopic equilibrium constants as denoted by superscripted 'i' and 'm', respectively. Panel D: ΔG formulation illustrating the relationship between microscopic and macroscopic ΔGs. Note in panels C and D that PL represents PL+LP from panels A and B. The Ks and the ΔGs are related by $\Delta G = -RT\ \ln K$.

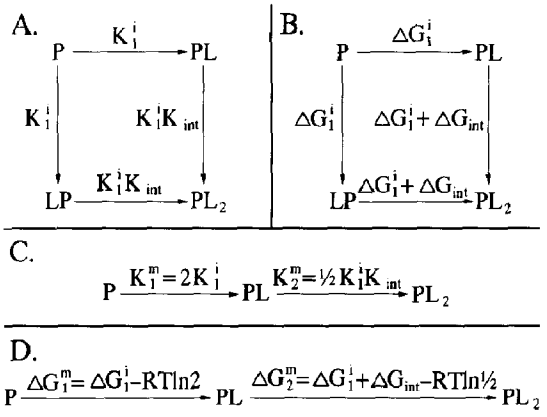


Fig. 2. Equivalent thermodynamic formulations for a homotropic two site protein incorporating the concept of interaction terms. Panels as in Fig. 1.

biochemical systems. A cooperative interaction is where an event at one site on a protein exerts an effect on events at other sites on the protein. The simplest system in which cooperative interactions can be discussed is for a homotropic two site protein as diagrammed in Fig. 1. If the binding of the first ligand has an effect on the second ligand's binding then the intrinsic ΔGs of binding, ΔG_1^i and ΔG_2^i, will not be equal. If ΔG_2^i is decreased (i.e. more negative; tighter binding) relative to ΔG_1^i then we have positive cooperativity and if it is increased (i.e. more positive; weaker binding) this is negative cooperativity. Note in Fig. 1 that if there is no cooperative interaction ΔG_1^i and ΔG_2^i are equal. If we consider that ΔG_2^i is in a sense the same as ΔG_1^i plus or minus some perturbation we can reformulate ΔG_2^i as

$$\Delta G_2^i = \Delta G_1^i + \Delta G_{int} \tag{6}$$

where ΔG_{int} is the ΔG of interaction. Using this definition we obtain the equivalent representations shown in Fig. 2. This reformulation results in two independent parameters (ΔG_1^i and ΔG_{int}). The ΔG_{int} is the actual amount of cooperative interaction expressed as a defined thermodynamic quantity. A positive ΔG_{int} means negative cooperativity and vice versa. This formulation is well known [9,10] and has been applied to the study of many systems.

Now consider a protein with two binding sites

each for a different ligand, A and B, as depicted in Fig. 3. In this scheme Weber's formulation is adopted where $\Delta G_{(A/B)}$ represents the ΔG of B binding to its site when the A site is already occupied [10]. We can reparameterize by noting that $\Delta G_{(A/B)}$ is the same as the ΔG of binding of B, b, plus a perturbation, ab. Also note that because of the conservation of energy requirement both pathways from P to PAB must have equal sums. Note that ab must appear as the interaction term on both pathways (i.e. both pathways must have equal sums) and this reformulation reflects the fact that whatever effect A binding has on B binding must be reciprocated. Our original scheme with four non-unique parameters is now reduced to a uniquely defined three parameter scheme.

Figure 4 illustrates the necessary thermodynamic formulation for three heterotropic interacting sites. Note that all terms are written starting from the free protein in the direction of increasing complexity. We wish to find terms which when substituted for Ia, Ib, and Ic will satisfy the requirement that all pathways leading between any two states must have equal sums. Starting from P and selecting pathways passing through the three terms of interest to PABC we get

$$a + b + c + ab + Ic = a + b + c + bc + Ia = a + b + c + ac + Ib \quad (7)$$

Canceling common terms gives

$$bc + Ia = ac + Ib = ab + Ic \quad (8)$$

One solution to this set of equations is if

$$Ia = ab + ac, \quad Ib = ab + bc, \quad \text{and} \quad Ic = ac + bc \quad (9)$$

but six terms cannot uniquely relate the seven complexes in this model to the standard state of the free components; seven terms are required. A unique solution can be obtained by adding another parameter, abc.

$$Ia = ab + ac + abc \quad (10)$$

$$Ib = ab + bc + abc \quad (11)$$

$$Ic = ac + bc + abc \quad (12)$$

Substituting back into the scheme gives Fig. 5. Note that, for example C binding to PAB, the ΔG of binding appears as a combination of terms equal to the unperturbed binding of C to P (c), two terms of interaction of C with the ligands already on the protein ($ac + bc$), and a third term common to all the pathways leading to the triply ligated species (the third order interaction term; abc). This reparameterization provides the minimal number of terms which uniquely defines each

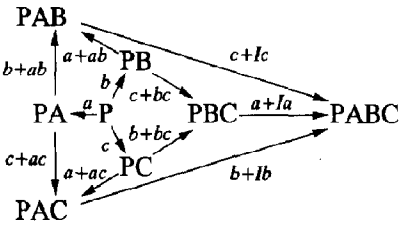


Fig. 4. Incomplete thermodynamic formulation for a heterotropic three site protein using interaction terms and the abbreviated formulation. The terms Ia, Ib, and Ic must be replaced with other terms which are derived in the text. As formulated here the model is unnecessarily overdetermined, since the seven complexes (relative to the free components) are described by nine terms; only seven terms are necessary to uniquely define the seven complexes.

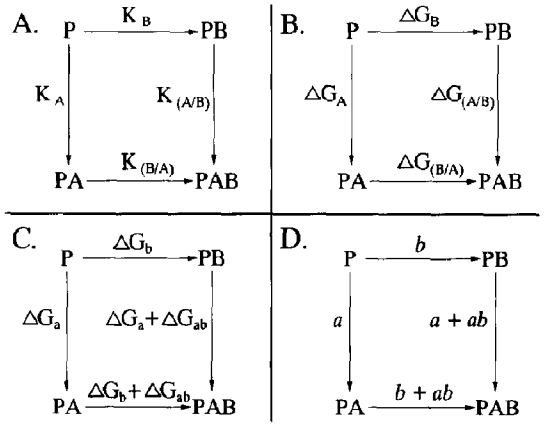
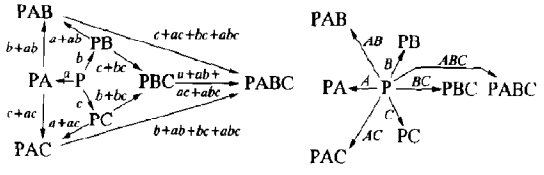


Fig. 3. Equivalent thermodynamic formulations for a heterotropic two site protein. Panel A: equilibrium constant formulation using conditional equilibrium constant formulation [10]. Panel B: equivalent ΔG formulation. Panel C: redefined ΔG formulation incorporating concept of interaction terms. Panel D: equivalent formulation introducing the abbreviated formulation where italicized lower case letters represent UIP formulation ΔGs.

species in the above scheme and still satisfies the requirement that the ΔG between any two species must be path independent. Reinhart appears to have first introduced the concept of a higher than second order interaction term in his analysis of allosterically regulated enzymes [14].

2.3 Hierarchical nature of the formulation

How does one add to an already existing model the presence of another ligand binding site? In Fig. 6 is shown a representation of Fig. 5 which illustrates the effect of the third ligand, C, on the A and B ligand binding pattern. The term *abc* can be seen to represent the difference in interaction between A and B with C bound ($ab + abc$) as opposed to unbound (ab) exactly in the same sense that ab represents the difference in the protein binding A with B bound ($a + ab$) as opposed to unbound (a). This illustrates the hierarchical nature of this formulation which should prove useful in those situations where an allosteric modifier for a protein ligand system is discovered after initial work on a system has been conducted. This allows for the analysis of the effect of the new ligand within the previously established framework (i.e. the discovery of 2,3-diphosphoglycerate as an allosteric modifier of



$a = A$
$b = B$
$c = C$
$ab = AB-A-B$
$ac = AC-A-C$
$bc = BC-B-C$
$abc = ABC-AB-AC-BC+A+B+C$

$A = a$
$B = b$
$C = c$
$AB = a+b+ab$
$AC = a+c+ac$
$BC = b+c+bc$
$ABC = a+b+c+ab+ac+bc+abc$

Fig. 5. Above left: reformulated heterotropic three site protein from Fig. 4. The model is uniquely defined by seven terms for the seven complexes. Above right: equivalent thermodynamic formulation where the capitalized italicized terms represent the total ΔG of formulation for a given complex relative to the free components. Below: interrelationships between the UIP formulation as shown on the above left and the total ΔG formulation on the above right illustrating the one-to-one correspondence between the formulations.

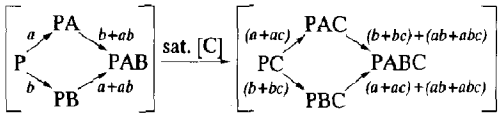


Fig. 6. Same model as Fig. 5 but from the perspective of illustrating the effect of a saturating (sat.) third ligand (C) on the observed thermodynamic model of A and B interactions. Note the similarity in the patterns between the left and right panels. Also note that with C bound our observed terms are all in a sense incremented by an interaction term with C (as if terms on the left are multiplied by a factor of $1 + c$ to obtain terms on the right). This figure illustrates that the presence of a third ligand, whether known or not, does not change the pattern of terms for the other two ligands, but does change the apparent values for the primary and second order terms as shown. This illustrates the hierarchical nature of the UIP formulation and its flexibility in handling complex system for which all the potential allosterically modulating ligands may not yet be known.

hemoglobin [15]). It also allows phenomena such as the Root effect to be explained within the context of a thermodynamic model [16] and provides insight into the meaning of the higher order terms. (The Root effect is the decrease in cooperativity observed in the hemoglobin from teleost fish upon going from a higher to lower pH.) Notice that what we are looking at in these diagrams can be considered slices of the complete model in much the same fashion as a plane of atoms in a crystal. However, our models have as many dimensions as ligand binding sites. We can subdivide the model into simpler "planes" by freezing out "dimensions". Similarly we can add new "dimensions" relatively easily when required.

2.4 Equilibrium constant equivalents of the interaction terms

There is a facile conversion between the numerical values of the ΔGs and equilibrium constants in the thermodynamic models discussed above. Equilibrium association constants are also defined symbolically as the ratio of product to reactant concentrations (assuming ideal behavior) for a given reaction. Referring to Fig. 5 we can define the equilibrium constant equivalents for all of the ΔGs in the UIP formulated model as

follows using the fact that addition of ΔGs corresponds to multiplication of equilibrium constants.

$$K_a = K_A = \frac{[PA]}{[P][A]} \tag{13}$$

$$K_{ab} = \frac{K_{AB}}{K_A K_B} = \frac{[PAB]}{[P][A][B]} \cdot \frac{[P][A]}{[PA]} \cdot \frac{[P][B]}{[PB]} = \frac{[PAB][P]}{[PA][PB]} \tag{14}$$

$$K_{abc} = K_{ABC} \cdot \frac{K_A K_B K_C}{K_{AB} K_{AC} K_{CB}} = \frac{[PABC]}{[P][A][B][C]} \cdot \frac{[PA]}{[P][A]} \cdot \frac{[PB]}{[P][B]} \cdot \frac{[PC]}{[P][C]} \cdot \frac{[P][A][B]}{[PAB]} \cdot \frac{[P][A][C]}{[PAC]} \cdot \frac{[P][B][C]}{[PBC]} = \frac{[PABC][PA][PB][PC]}{[PAB][PAC][PBC][P]} \tag{15}$$

As a final point, consider the homotropic two site model from Fig. 2 as defined using the abbreviated formulation

$$P \xrightarrow{l - RT\ln(2)} PL \xrightarrow{l + ll - RT\ln(\frac{1}{2})} PL_2 \tag{16}$$

The total ΔGs for PL and PL_2 are, respectively

$$L = l - RT\ln(2) \tag{17}$$

$$LL = l - RT\ln(2) + l + ll - RT\ln(\tfrac{1}{2}) = 2l + ll \tag{18}$$

By definition the equilibrium constant equivalents for L and LL are

$$K_L = \frac{[PL]}{[P][L]} \tag{19}$$

$$K_{LL} = \frac{[PL_2]}{[P][L]^2} \tag{20}$$

From these relationships the equilibrium constant equivalents for l and ll can be derived as

$$K_l = \frac{[PL]}{[P][L]} \cdot \frac{1}{2} \tag{21}$$

$$K_{ll} = \frac{[PL_2][P]}{[PL]^2} \cdot 4 \tag{22}$$

The constants appearing in eqs. (21) and (22) are important in maintaining the internal consistency of UIP equilibrium constant formulated schemes for theoretical applications of the reformulation as shown in the following paper.

2.5 Physical significance of higher order interaction terms

The primary terms have an obvious physical significance representing the ΔG (or K) of ligand binding to an unperturbed site. The second order terms represent the mutual effect of one ligand on another; *ab* represents the effect of A binding on B binding and vice versa. The equivalent K_{ab} has a form which describes the equilibrium distribution of the ligands on the protein. If it is equal to unity ($ab = 0$; no cooperativity) then A and B will be distributed randomly among their respective binding sites. The higher order terms present more of a conceptual enigma. *abc* can be considered the mutual effect A has on *bc*, B has on *ac*, and C has on *ab*. The corresponding equilibrium constant, K_{abc}, is more complex than K_{ab} reflecting the more complex nature of the distribution of ligands associated with this term. One way to look at what physical interpretation can be ascribed to this term is to divide it into simpler constituents. One way to do this for K_{abc} is as follows

$$K_{abc} = \frac{[PABC][PA]}{[PAB][PAC]} \div \frac{[PBC][P]}{[PB][PC]} = \frac{K_{bc}\ (\text{with A already bound})}{K_{bc}} = \frac{[PABC][PB]}{[PAB][PBC]} \div \frac{[PAC][P]}{[PA][PC]} = \frac{K_{ac}\ (\text{with B already bound})}{K_{ac}} = \frac{[PABC][PC]}{[PAC][PBC]} \div \frac{[PAB][P]}{[PA][PB]} = \frac{K_{ab}\ (\text{with C already bound})}{K_{ab}} \tag{23}$$

This analysis illustrates that the higher order terms reflect the modulation of the lower order terms upon binding of ligands in a hierarchical fashion.

2.6 First set of generalized rules for constructing UIP defined schemes

The above formalization for obtaining UIP formulated schemes for multiple ligand binding schemes can be summarized in the following set of rules.

(1) To simplify the schemes for multiple identical binding sites the macroscopic ΔG of ligand binding is obtained in terms of the intrinsic ΔG of ligand binding as

$$\Delta G^{m} = \Delta G^{i} - RT \ln \frac{\text{No. of ways forward to product}}{\text{No. of ways back from product}} \tag{24}$$

or

$$\Delta G^{m} = \Delta G^{i} - RT \ln(n/t - n + 1) \tag{25}$$

where t denotes the total number of identical sites for the current ligand, and n is the number of unoccupied sites for the current ligand; i.e number of ways forward, and $t - n + 1$ is the number of ways back from the product of the step. For example, in a three identical site protein the macroscopic ΔGs are (ignoring interaction terms)

$$\mathrm{P} \xrightarrow{l - RT \ln(\frac{3}{1})} \mathrm{PL} \xrightarrow{l - RT \ln(1)} \mathrm{PL_2} \xrightarrow{l - RT \ln(\frac{1}{3})} \mathrm{PL_3} \tag{26}$$

(2) Interaction terms are determined in a hierarchical fashion including increasingly higher order terms until the highest term, which is equal in order to the number of ligands bound to the product of the step. In doing this each possible assortment with ligands already on the protein must be included. As an example consider the scheme above. The ΔG^{m} for the first ligand binding will be, l, plus the statistical term, $-RT$ ln(3). The ΔG^{m} for the second ligand binding will be equal to the primary term, l, plus a second order interaction term ll, plus the statistical correction which in this case is 0. The ΔG^{m} for the third ligand binding will contain the base term, l, plus the second order interaction terms with each of the already bound ligands, $ll + ll$, plus the triple interaction term, lll, plus the statistical correction which in this case is $-RT \ln(\frac{1}{3})$. The complete scheme will be;

$$\mathrm{P} \xrightarrow{l - RT \ln(3)} \mathrm{PL} \xrightarrow{l + ll} \mathrm{PL_2} \xrightarrow{l + ll + ll + lll - RT \ln(\frac{1}{3})} \mathrm{PL_3} \tag{27}$$

Note that the number of terms, three (l, ll, and lll), is exactly equal to the number of complexes in the system and therefore represents the minimal number necessary to "uniquely" relate each complex to the "standard state" of free protein and free ligand. Also note that all the interaction terms will be zero if there is no interaction between binding events. This parameterization is not the only one available. A ΔG could simply be assigned to each complex relative to the free components as illustrated in Fig. 5. The purpose of the reparameterization is twofold. First, it provides a set of parameters which will be uniquely defined and independent whether or not any cooperative interaction is present. Second, it provides a formulation based upon the inherent symmetry of the system, the intrinsic properties of the individual binding sites, and the potential hierarchy of interactions between them.

2.7 Example derivations of microscopic and macroscopic models of hemoglobin oxygen binding

Consider as an example hemoglobin with two α and two β chains. The construction of the model starts from the free components. In a thermodynamic sense these are our standard states. Note that if one is dealing with a protein which can oligomerize then there will be the option of formulating with the free subunits as the standard state, or to ignore the oligomerization equilibrium and start with the oligomer as the standard state. In this analysis oligomerization equilibria will be ignored and the tetrameric state will be the standard state. Hemoglobin will

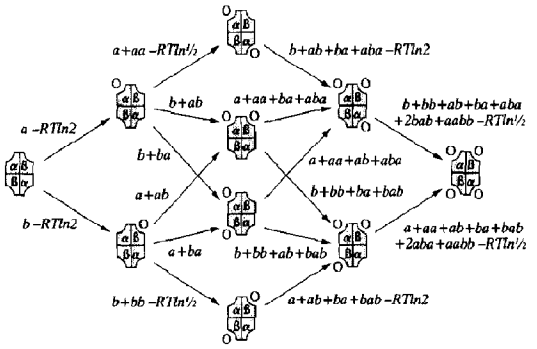


Fig. 7. A microscopic model of oxygen binding to hemoglobin with discrimination between the α and β chains to illustrate the derivation of terms for sequential ligand binding in a complex system. Note that the number of terms is equal to the number of complexes and that the total ΔG between any two species is independent of path.

be represented to distinguish between the α and β chains.

The scheme in Fig. 7 can be derived. There are several assumptions implicit in this model. One is that the interaction terms *ab* and *ba* are different as might be expected based upon the asymmetry of the protein [1]. Another is that we can distinguish each of the binding events in terms of the individual species in the mixture. Notice that for the complete description of even moderately complex schemes a large number of terms appear. It is not expected that all of these will be determinable, at least at the outset of an experimental investigation. If no distinction is possible between the binding sites or of the geometric relationship of the bound ligands in the complexes the appropriate model would be

$$\mathrm{Hb} \xrightarrow{o - RT\ln 4} \mathrm{HbO} \xrightarrow{o + oo - RT\ln(\frac{3}{2})} \mathrm{HbO_2}$$
$$\xrightarrow{o + 2oo + ooo - RT\ln(\frac{2}{3})} \mathrm{HbO_3}$$
$$\xrightarrow{o + 3oo + 3ooo + oooo - RT\ln(\frac{1}{4})} \mathrm{HbO_4} \qquad (28)$$

where o is the primary ΔG of binding, oo is the second order interaction term, ooo is the third order interaction term, and $oooo$ is the fourth order interaction term. The total ΔGs of formation for each complex in the above model from components are

$$\Delta G_{\mathrm{HbO}} = o - RT\ln(4) = o - RT\ln 4 \qquad (29)$$

$$\Delta G_{\mathrm{HbO_2}} = \Delta G_{\mathrm{HbO}} + o + oo - RT\ln(\tfrac{3}{2})$$
$$= 2o + oo - RT\ln 6 \qquad (30)$$

$$\Delta G_{\mathrm{HbO_3}} = \Delta G_{\mathrm{HbO_2}} + o + 2oo + ooo - RT\ln(\tfrac{2}{3})$$
$$= 3o + 3oo + ooo - RT\ln 4 \qquad (31)$$

$$\Delta G_{\mathrm{HbO_4}}$$
$$= \Delta G_{\mathrm{HbO_3}} + o + 3oo + 3ooo + oooo - RT\ln(\tfrac{1}{4})$$
$$= 4o + 6oo + 4ooo + oooo \qquad (32)$$

2.8 Second set of rules for constructing UIP defined schemes

It is now possible to state a straightforward method for determining what the appropriate term will be for either a single step in a complex protein ligand binding scheme or for the total ΔG of formation of any complex by inspection. Note that the pattern which emerges above for hemoglobin is that the association of the new ligand at a particular step is represented by a term equal to the primary term, plus terms for combinations of the new ligand with previously bound ligands, minus the statistical correction term for the step. Similarly the total ΔG of formation of a given complex from the standard state is represented by a set of terms equal to all possible associations of the ligands, starting with one ligand taken at a time and increasing stepwise up to the total number of ligands, minus a statistical correction term equal to the number of ways the complex can be described. For example consider the total ΔG of formation of the triply ligated hemoglobin species as shown in Fig. 7 with two α subunits and one β subunit ligated to oxygen. First we consider all base terms, so we have

$$2a + b \qquad (33)$$

Then we consider the second order terms

$$aa + ab + ba \qquad (34)$$

Next we have the third order term

$$aba \tag{35}$$

Finally we need the statistical correction term. In this case there are two ways of describing the complex, with the β oxygen on either of the two β subunits so we get

$$-RT \ln 2 \tag{36}$$

All these together give

$$2a + b + aa + ab + ba + aba - RT \ln 2 \tag{37}$$

This result is easily validated by adding up terms along the pathways leading from the free protein to this complex shown in Fig. 7.

3. Conclusions

We describe here a method for reformulating protein ligand binding equilibria in terms of a set of unique and independent ΔGs. This reformulation recasts the N ΔGs necessary to completely define a model with N complexes (a complex being defined as any states other than the standard state) into a new set of N ΔGs. These new ΔGs are composed of a set of base (primary) terms which describe the unperturbed ΔG of ligand binding and a hierarchy of higher order terms reflecting possible interactions between the binding of ligands. The concept of interaction energy, as shown in Figs. 2B and 3C, was first proposed by Wyman [9] and developed additionally by Weber [10]. That more complex models than two ligands (conformations) require more than second order interaction terms appears to have gone essentially unnoticed except for the analysis of allosterically regulated enzymes by Reinhart [14] where the concept of third order interaction term was introduced. The purpose of the analysis presented here is to extend the concept of interaction energy to the general case of N ligands. In the following paper this formalism is extended to include conformational transitions as well as ligand binding [11]. This formalism has a number of important applications ranging from the theoretical analysis of the thermodynamics of complex systems, as demonstrated in the following paper, to the analysis of equilibrium binding data pertaining to complex systems [17]. This reformulation is not specific to biochemical systems but is generally applicable to any physical system at equilibrium. Given their intrinsic complexity the analysis is particularly pertinent to biochemical systems.

Acknowledgement

This work was supported by NIH Grant GM31716 and NIH postdoctoral fellowship 1F32HL08314-01 (WGG).

References

1 M.F. Perutz, Q. Rev. Biophys. 22 (1989) 139–236.
2 W.P. Jencks, Advances in enzymology and related areas of molecular biology, 43 (Wiley, New York, 1975) pp. 219–402.
3 J. Monod, J. Wyman and J.P. Changeux, J. Mol. Biol. 12 (1965) 88–118.
4 D.E. Koshland, Jr., G. Nemety and D. Filmer, Biochemistry 5 (1966) 365–385.
5 M. Eigen, Nobel Symp. 5 (1967) 333–369.
6 K.E. Van Holde, (1971) Physical biochemistry (Prentice-Hall, Englewood Cliffs, NJ, 1977) pp. 57–62.
7 S.J. Gill, E. Di Cera, M.L. Doyle, G.A. Bishop and C.H. Robert, Biochemistry 26 (1988) 3995–4002.
8 K. Imai, Biophys. Chem. 37 (1990) 197–210.
9 J. Wyman, Adv. Protein Chem. 18 (1964) 223–286.
10 G. Weber, Adv. Protein Chem. 29 (1975) 1–83.
11 W.G. Gutheil, Biophys Chem., 45 (1992) 181–191.
12 G.D. Reinhart, S.B. Hartleip and M.M. Symcox, Proc. Natl. Acad. Sci. USA, 86 (1989) 4032–4036.
13 C. Tanforth, in: Physical chemistry of macromolecules, Chapter 8, (Wiley, New York, 1961).
14 G.D. Reinhart, Arch. Biochem. Biophys. 224 (1983) 389–401.
15 R. Benesch, R.E. Benesch and C.I. Yu, Proc. Natl. Acad. Sci. USA 59 (1968) 526–532.
16 M. Brunori, Curr. Top. Cell. Regul. 9 (1975) 1–39.
17 W.G. Gutheil, Ph.D. Dissertation, University of Southern California, Los Angeles, CA (1989).