

## A Simple Model for Proteins with Interacting Domains. Applications to Scanning Calorimetry Data<sup>†</sup>

John F. Brandts,\* Cui Qing Hu,<sup>‡</sup> and Lung-Nan Lin

*Department of Chemistry, University of Massachusetts, Amherst, Massachusetts 01003*

Maria T. Mas

*Physical Biochemistry Section, Beckmann Institute of the City of Hope, Duarte, California 91010*

*Received February 3, 1989; Revised Manuscript Received May 23, 1989*

**ABSTRACT:** A simple thermodynamic model is formulated for the purpose of interpreting scanning calorimetry data on proteins that have interacting domains. Interactions are quantified by inclusion of an interface free energy,  $\Delta G_{AB}$ , in the thermodynamics of unfolding for multidomain proteins. The assumption is made that  $\Delta G_{AB}$  goes to zero with the unfolding of either domain involved in pairwise interaction, so the interaction term appears to stabilize only the domain with the lower  $T_M$ . Application of the model to calorimetric data leads to an estimate of  $-25\,000$  cal/mol for interactions between the regulatory and catalytic subunits of native aspartate transcarbamoylase and to a value of 0 for  $\Delta G_{AB}$  between the transmembrane and cytoplasmic domains of band 3 of the human erythrocyte membrane. Estimates of changes in  $\Delta G_{AB}$  are also obtained for mutant forms of yeast phosphoglycerate kinase that have been altered in the hinge region between amino-terminal and carboxy-terminal domains. The model is also applied to ligand binding to proteins having domains that communicate through pairwise interaction. It is shown that whenever the  $\Delta G_{AB}$  term is ligand-dependent, then attachment of the ligand to the binding domain will be partially controlled by the other (regulatory) domain. This situation can sometimes be recognized and quantified when calorimetric scans are carried out at varying ligand concentrations. According to the model, the binding of MgATP to the carboxy-terminal domain of phosphoglycerate kinase is strongly stabilized (ca. 20% of the unitary free energy of binding) by participation of the amino-terminal domain, which acts to increase the binding constant 25-fold. A similar situation is found for binding of MgADP to creatine kinase, except the contribution from the regulatory domain is much smaller (ca. 5% of the unitary free energy of binding) and only increases the binding constant 2-fold.

One of the important structural elements of proteins is the domain (Edelman, 1970; Wetlaufer, 1973, 1981) or, as it is less commonly called, the cooperative unit. Recognition of the significance of domains has increased as evidence accumulates to show that single domains frequently comprise biological functional units (Edelman, 1970; Rao & Rossman, 1973; Janin, 1979; Rossman & Argus, 1981) and that domain-domain interactions provide the intraprotein communication that is necessary to coordinate function.

Differential scanning calorimetry (DSC) has frequently been used in the study of multidomain proteins since it allows individual domains to be seen as they undergo cooperative thermal unfolding. If the domain transitions are well separated on the temperature axis, then thermodynamic parameters  $\Delta H$  and  $T_M$  may be obtained directly for each domain. Even when transitions overlap severely in temperature, available software allows rapid deconvolution of the transition envelope to obtain estimates of the individual-domain parameters.

All deconvolutions are based on thermodynamic models, and there has in the past been a reluctance to apply these models to proteins known to unfold irreversibly or, worse, to precipitate toward the end or after completion of the transitions. However, there is now definitive evidence to show that many systems having little or no reversibility on the second heating do in fact conform closely to thermodynamic behavior during the short time they spend in the transition region during the

first heating. This less conservative view has been presented convincingly by Sturtevant (Sturtevant, 1987; Manly et al., 1985; Edge et al., 1985, 1988; Hu & Sturtevant, 1987) and is being cautiously adopted by many labs.

Of the two thermodynamic models available for the interpretation of DSC results on multidomain proteins, one (model of independent transitions) neglects domain interactions entirely while the other (model of sequential transitions) assumes implicitly that domain interactions are always strong enough to force an immutable order to the sequence in which domains must unfold. Neither of these models is capable of providing any quantitative information on the strength of domain interactions or how these interactions might change with conditions.

In this paper, a new model is described which incorporates domain interactions through a pairwise interaction free energy  $\Delta G_{AB}$  at the A-B domain interface. The model becomes operational by making the assumption that the structural integrity of both domains is necessary for the interactions to persist; i.e., the unfolding of either domain causes  $\Delta G_{AB}$  to go to 0. Although this treatment becomes equivalent to that of independent transitions when  $\Delta G_{AB}$  is 0 and to that of sequential transitions when  $\Delta G_{AB}$  is very large, the incorporation of an adjustable interaction term provides insights and opportunities not present in either of the limiting models.

Under the most favorable circumstances, interpretation of DSC data using the model will (1) provide numerical estimates of the total  $\Delta G_{AB}$  for interacting domains, (2) provide estimates of changes in  $\Delta G_{AB}$  brought about by changing solution conditions, introducing mutations, etc., and (3) recognize when ligand interactions with a binding domain are being influenced

<sup>†</sup> This project was supported by Grant GM-11071 from the National Institutes of Health.

<sup>‡</sup> Present Address: Institute of Biophysics, Academia Sinica, Peking, China.

by a nearby regulatory domain and provide numerical estimates of the increase (or decrease) in binding constant that is actually due to ligand-dependent changes in interactions between the binding domain and the regulatory domain. These applications are illustrated by using experimental DSC data.

## MATERIALS AND METHODS

Wild-type and mutant forms of yeast phosphoglycerate kinase (PGK) were cloned and expressed in PGK<sup>-</sup> *Saccharomyces cerevisiae* strain XSB44-35D according to procedures described elsewhere (Mas et al., 1987, 1988). Bovine RNase (catalog no. R-5500) and rabbit muscle creatine kinase (CK) (catalog no. C-3755) were purchased from Sigma Chemical Co. and used without further purification. All other chemicals were of reagent grade.

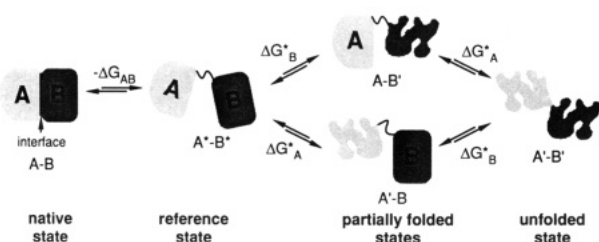
Concentrations of all protein samples were determined spectrophotometrically by using the following extinction coefficients: RNase, 9800 M<sup>-1</sup> cm<sup>-1</sup> at 278 nm; CK, 0.888 mL mg<sup>-1</sup> cm<sup>-1</sup> at 280 nm; PGK, 0.49 mL mg<sup>-1</sup> cm<sup>-1</sup> at 280 nm.

Calorimetric measurements were carried out on a MicroCal MC-2 ultrasensitive DSC (MicroCal Inc., Northampton, MA) using the standard DA-2 software package for data acquisition, analysis, and deconvolution. Solutions of PGK were quick frozen, stored at -80 °C, and thawed immediately before use. Solutions of RNase and CK were used within 24 h after preparation. All protein solutions and buffers were degassed ca. 1 min with gentle stirring under vacuum (27 mmHg) before being loaded into the calorimeter; the vacuum filling device was used to prevent bubble entrapment in the total-fill cells. Buffer-buffer or water-water base lines were subtracted from sample data before normalization on the scan rate.

## MODEL OF INTERACTING DOMAINS

### Model Formulation and Simulations

The states that are assumed to exist for the simplest case, a two-domain protein, are



where A and B signify folded domains and A' and B' signify unfolded domains. The domains are illustrated here as being in the same polypeptide chain, but could also be two interacting subunits in different chains. In the native state A-B, interactions between the two domains are quantified by the interface term,  $\Delta G_{AB}$  (with corresponding  $\Delta H_{AB}$  and  $\Delta S_{AB}$ ). The reference state A'-B' corresponds to separated domains where both A and B are still folded but not interacting ( $\Delta G_{AB} = 0$ ). This reference state will not normally exist at significant concentration in the high-temperature region where the native protein melts since its formation will be quickly followed by unfolding of one or both domains. It may be possible in some cases to isolate the reference state experimentally under more stabilizing conditions by separation of intact-chain subunits, by gentle proteolysis and isolation of multiple domains contained within the same chain, by introducing mutations that critically destabilize the AB interface, or by cloning abbreviated sequences that comprise only one domain. In other cases, the stability of folded domains in the absence of interface

interactions may be too low to permit their isolation under any conditions. The reference state in these latter cases will be a hypothetical state which acts as a convenient intermediate between the initial and final states in the unfolding scheme. The standard free energies for unfolding domains in the reference state are  $\Delta G^*_A$  and  $\Delta G^*_B$ , transition midpoints are  $T^*_A$  and  $T^*_B$ , and enthalpy changes are  $\Delta H^*_A$  and  $\Delta H^*_B$ . For the moment, it will be assumed that all enthalpies are temperature-independent and all single-domain transitions are of the two-state type.

In practice,  $\Delta G_{AB}$  is an empirically defined term and will include, in addition to interactions directly at the interface, conformational changes in A or B that occur as a result of the interface. Although  $\Delta G_{AB}$  is expected to be negative for most systems, this does not have to be the case. An important feature of this model, dictated by the placement of the reference state prior to the partially unfolded states in the reaction scheme, is the assumption that the interactions at the interface only exist if both domains are in their low-temperature native state. This means that the  $\Delta G_{AB}$  term must go to 0 whenever either domain melts, so that *domain-domain pairwise interaction appears to affect only the domain with the lower  $T_M$* . That some multidomain proteins conform closely to this prediction of the model will be shown later.

With these assumptions, the model may be treated exactly by using conventional thermodynamic methods. The free energy of each of the states A'-B, A-B', and A'-B' (relative to a value of 0 for the native A-B state at each temperature) and their corresponding fractional populations determined from these free energies are

$$\Delta G_{A'-B} = \Delta G^*_A - \Delta G_{AB} = \Delta H^*_A(1 - T/T^*_A) - \Delta G_{AB} = -RT \ln C_1$$

$$\Delta G_{A-B'} = \Delta G^*_B - \Delta G_{AB} = \Delta H^*_B(1 - T/T^*_B) - \Delta G_{AB} = -RT \ln C_2$$

$$\Delta G_{A'-B'} = \Delta G^*_A + \Delta G^*_B - \Delta G_{AB} = \Delta H^*_A(1 - T/T^*_A) + \Delta H^*_B(1 - T/T^*_B) - \Delta G_{AB} = -RT \ln C_3$$

$$C_{\text{total}} = 1 + C_1 + C_2 + C_3$$

$$f_{A-B} = 1/C_{\text{total}} \quad f_{A'-B} = C_1/C_{\text{total}}$$

$$f_{A-B'} = C_2/C_{\text{total}} \quad f_{A'-B'} = C_3/C_{\text{total}}$$

$$H(T) = f_{A-B}(\Delta H^*_A - \Delta H_{AB}) + f_{A-B'}(\Delta H^*_B - \Delta H_{AB}) + f_{A'-B'}(\Delta H^*_A + \Delta H^*_B - \Delta H_{AB})$$

$$C_p(T) = dH(T)/dT$$

Once the parameters  $\Delta H^*_A$ ,  $\Delta H^*_B$ ,  $T^*_A$ ,  $T^*_B$ ,  $\Delta G_{AB}$ , and  $\Delta H_{AB}$  are assigned, the system is defined so that the fraction  $f$  of each species may be calculated as a function of temperature, as may the system enthalpy  $H(T)$ . This allows the excess heat capacity  $C_p(T)$  to be determined so that DSC curves may be simulated for any values of input parameters.

Some simulated curves are shown in Figure 1 for a hypothetical two-domain protein, each domain having a  $\Delta H^*$  of 100 kcal but with differing intrinsic transition midpoints  $T^*_A$  and  $T^*_B$  as indicated in the upper scan. The interaction term  $\Delta G_{AB}$  was chosen as -2000 cal and was entirely enthalpic. Since the A domain has the lowest intrinsic midpoint, its experimental midpoint  $T_A$  is shifted higher by domain interactions according to the equation

$$\Delta G_{AB} = -\Delta H_A(1 - T^*_A/T_A) + \Delta C_p(T^*_A \ln [T^*_A/T_A] + T_A - T^*_A) \quad (1)$$

and when  $\Delta C_p$  is 0,<sup>1</sup> this becomes

$$\Delta G_{AB} = -\Delta H_A(1 - T^*_A/T_A) \quad (2)$$

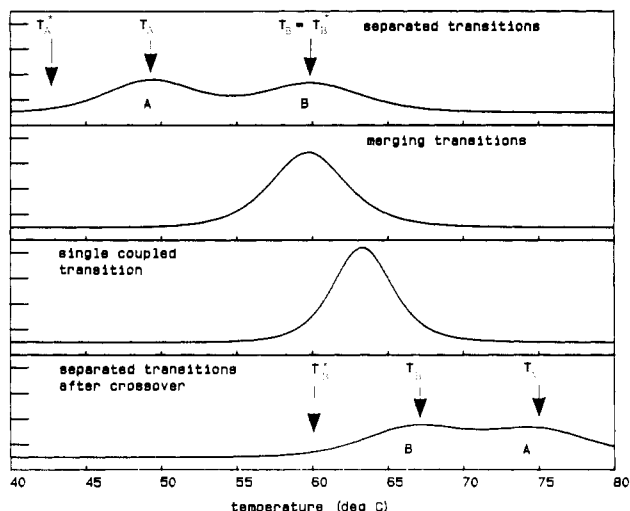


FIGURE 1: Simulated DSC scans for a two-domain protein ( $\Delta H_A^* = \Delta H_B^* = 100$  kcal/mol) with  $\Delta G_{AB}$  of  $-2000$  cal. The value of  $T_B^*$  is  $60^\circ\text{C}$  for all simulations, while  $T_A^*$  varies from  $43$  to  $53$  to  $60$  to  $75^\circ\text{C}$  in going from the top frame to the bottom frame.

where  $\Delta G_{AB}$  is for the temperature  $T_A^*$  [the equivalent expression at  $T_A$  is  $\Delta G_{AB} = \Delta H_A^*(1 - T_A/T_A^*)$ ]. The B transition occurs at  $T_B$ , which is identical with its intrinsic  $T_B^*$  since the interface interactions have already disappeared during the A transition and cannot act to stabilize B. Whenever  $T_A$  and  $T_B$  are well separated and transition overlap is minimal, deconvolution of the data shows both transitions to be independent two-state processes and no information is gained as to whether A and B are interacting or noninteracting domains (unless  $T_A^*$  were known from other experiments, as discussed later for aspartate transcarbamoylase). It is possible to differentiate between these two situations if the lower transition can be shifted higher (e.g., by adding a high concentration of a ligand that binds to the A domain only, thereby making  $\Delta G_A^*$  more positive) or if the higher transition can be shifted lower (e.g., by introducing a destabilizing mutation in the B domain to decrease  $\Delta G_B^*$ ) to produce either merging or crossover of  $T_A$  and  $T_B$ . The bottom three scans in Figure 1 depict the sequence of events as  $T_A^*$  is progressively increased, with  $T_B^*$  and other parameters including  $\Delta G_{AB}$  being held constant. As the two transitions begin to merge (second scan), the transition envelope can no longer be fit by two overlapping independent transitions (although two sequential transitions will fit) since the interdomain cooperativity caused by the  $\Delta G_{AB}$  term adds increased sharpness to the profile. As  $T_A^*$  is further increased (third scan), the two transitions couple even more strongly and the transition envelope deconvolutes as a single two-state transition (i.e., if  $\Delta G_{AB}$  is large enough,

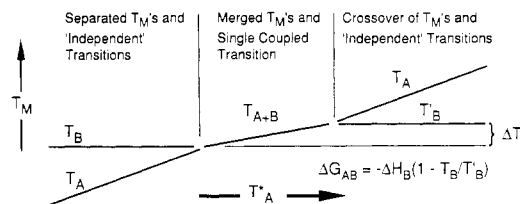


FIGURE 2: Schematic illustration of the variations of  $T_A$  and  $T_B$  that are brought about by changing  $T_A^*$  while keeping all other model parameters constant. See text for details.

as discussed below) rather than as two overlapping independent transitions. Further increases in  $T_A^*$  will continue to shift this coupled A + B transition, but the point will eventually be reached (last scan) where the transitions again uncouple and the A transition completely crosses over the B transition. The midpoint  $T_B$  for the B transition will be higher after crossover than before merging since it is now stabilized by the domain interaction term.

The entire process of merging, coupling, and crossover is illustrated schematically in Figure 2, which shows the two regions where separated independent transitions occur and the central region where the two domains couple into a single transition. The temperature span encompassed by the central region depends on the magnitude of  $\Delta G_{AB}$ . If this interaction term is very small, then the two transitions will not couple at all and crossover occurs very abruptly with  $T_B'$  equal or nearly equal to  $T_B$ . Very strong domain interactions lead to a large temperature interval for the coupled transition, and in the extreme, some proteins may not exhibit DSC evidence for domain substructure under any achievable conditions. In those cases where complete crossover occurs, then the absolute value of the interaction term can be readily obtained at  $T_B'$  from the equation

$$\Delta G_{AB} = \Delta H_B(1 - T_B'/T_B) + \Delta C_p(T_B' \ln [T_B/T_B'] + T_B' - T_B) \quad (3)$$

or at  $T_B$  from the equation

$$\Delta G_{AB} = -\Delta H_B'(1 - T_B/T_B') + \Delta C_p(T_B \ln [T_B/T_B'] + T_B - T_B') \quad (4)$$

where  $\Delta H_B$  is the enthalpy change at transition midpoint  $T_B$ ,  $\Delta H_B'$  is the corresponding parameter at midpoint  $T_B'$ , and  $\Delta C_p$  is the temperature-independent heat capacity change for the transition. If  $\Delta C_p$  is 0, as in these simulations, then only the first term on the right-hand side of eq 3 and 4 is necessary. When coupling can be achieved, but not complete crossover, then an estimate of the lower limit of  $\Delta G_{AB}$  can be made from the same equation by replacing  $T_B'$  by the maximum value of  $T_{A+B}$  that can be achieved.

Simulations show that, if two domains have exactly the same intrinsic stability ( $T_A^* = T_B^*$ ), then even weak interactions are sufficient to couple the DSC transitions. If the ratio  $\Delta H_{cal}/\Delta H_{vh}$  for the transition envelope is used as the coupling index, then this index will change ca. 90% of its maximum as  $\Delta G_{AB}$  is changed from 0 to  $-2500$  cal (e.g., for two superimposed transitions of equal  $\Delta H$ , the coupling index has a value of 2.0 if  $\Delta G_{AB}$  is 0, 1.0 for infinitely strong interaction, and 1.1 when  $\Delta G_{AB}$  is  $-2500$  cal, and this is independent of the magnitude of the individual  $\Delta H$ 's).

For the more usual case when  $T_A^*$  is substantially lower than  $T_B^*$ , the magnitude of  $\Delta G_{AB}$  necessary to couple into a single transition will be approximately equal to the sum of that required to superimpose the two transitions plus that required to couple the superimposed transitions, i.e.

<sup>1</sup> Even though the model simulations assume a zero heat capacity of unfolding for simplicity, most of the derived equations assume a temperature-independent  $\Delta C_p$  and are therefore more appropriate for real experimental systems. Free energy extrapolations from a transition midpoint  $T_M$  to a final temperature  $T$  can be made from the equation (Becktel & Schellman, 1987)  $\Delta G(T) = \Delta H_M(1 - T/T_M) - \Delta C_p([T_M - T] + T \ln [T/T_M])$ , where  $\Delta G(T)$  is the free energy at  $T$ ,  $\Delta H_M$  is the enthalpy change for unfolding at  $T_M$ , and  $\Delta C_p$  is the temperature-independent change in heat capacity for unfolding. In those cases where the extrapolation from  $T_M$  to  $T$  is fairly short, the inclusion of the heat capacity term in the extrapolation is unnecessary since its effect is smaller than the error introduced into the extrapolation from experimental uncertainty in  $\Delta H_M$ . However, when the extrapolation exceeds ca.  $5^\circ\text{C}$ , then neglect of the heat capacity term may introduce additional uncertainties in  $\Delta G(T)$ . These uncertainties become exaggerated at low temperatures where  $\Delta H$  becomes small in magnitude (Becktel & Schellman, 1987).

$$-\Delta G_{AB} = 2500 + \Delta H^*_A(T^*_B/T^*_A - 1) \quad (5)$$

when heat capacity effects are neglected. Since the second term on the right-hand side of eq 5 can be very much larger than the first term, some "multidomain proteins" with relatively weak seams could appear to be single-transition proteins in DSC scans while others with strong seams could have multiple, independent transitions, the exact behavior depending critically on differences in intrinsic stability of the domains as reflected by the relative values of  $T^*_A$  and  $T^*_B$ . DSC results may then at times appear contradictory with visual impressions formed from X-ray crystal structures, which provide some qualitative feeling for domain interactions but no sense of differences in intrinsic domain stabilities. This may help to explain the recent statement of Hu and Sturtevant (1987): "We have studied by means of DSC three proteins that according to X-ray crystallographic data have bilobate structures, namely, yeast hexokinase, the arabinose binding protein of *E. coli*, and phosphoglycerate kinase. Of these three only the first gave definite DSC indication of the more or less independent unfolding of the two lobes, and this only in the absence of substrate."

### Comparison with Experimental Results

**Aspartate Transcarbamoylase (ATCase).** Native ATCase exists as a tight complex of six regulatory subunits  $r$  and six catalytic subunits  $c$  to form  $c_6r_6$ . DSC experiments have been carried out (Edge et al., 1988) under identical conditions on native ATCase and on both the isolated regulatory subunit (which normally exists as a dimer,  $r_2$ ) and the isolated catalytic subunit ( $c_3$ ). The thermal midpoint for the regulatory subunit (corresponding to the A domain in the model) increases by 16 deg from 51 °C for the isolated subunit to 67 °C in native ATCase. The thermal midpoint for the higher melting catalytic domain remains unchanged at 73 °C in the isolated subunit and in native ATCase, in keeping with the model prediction. Although the situation for this protein is a little more complicated than for the model protein discussed above (since two subdomains of nearly equal stability exist in the regulatory subunits and three subdomains exist in catalytic subunits both in the native protein and the isolated subunits), the overall interaction free energy  $\Delta G_{AB}$  between regulatory and catalytic subunits in native ATCase may be estimated from eq 2 by considering each subdomain separately. Proceeding in this manner, the numerical value for  $\Delta G_{AB}$  at 51 °C, obtained from the data in Table I of Edge et al. (1988), is -25 000 cal per mole of native ATCase, or about -4200 cal of interaction for each catalytic and regulatory subunit.<sup>2</sup> This estimate neglects heat capacity effects since  $\Delta C_p$  values were not obtained by Edge et al.

As would be expected, there is also a large increase in  $\Delta H$  for the regulatory subunit (80% higher in native ATCase at 67 °C than in the isolated subunit at 51 °C) and almost no change (3% higher in the native form) for the catalytic subunit. This leads to a  $\Delta H_{AB}$  of -220 000 cal per mole of native AT-

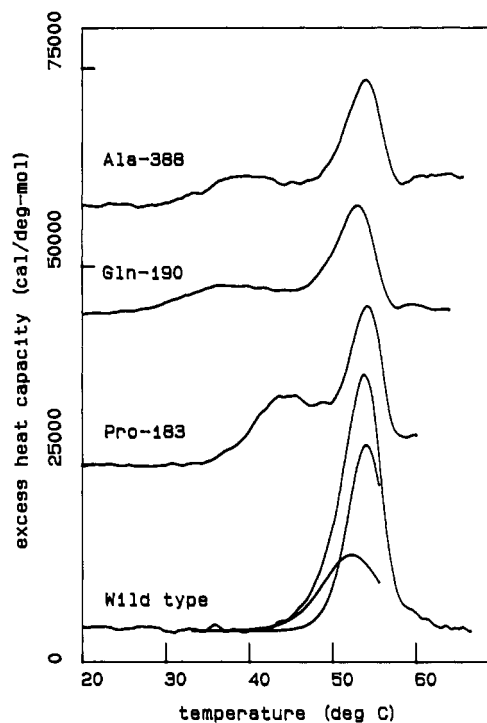


FIGURE 3: Experimental DSC scans on wild-type yeast phosphoglycerate kinase and on three mutant forms modified near the hinge region. For the wild-type PGK, the two curves below the experimental curve show the two individual overlapping transitions obtained by deconvolution of the experimental curve. For the three mutants, the individual transitions for the amino- and carboxy-terminal domains are well separated due to a mutation-induced shift downward in  $T_M$  for the transition of the amino-terminal domain, so that convolution results are not shown. Solution conditions were identical for all scans (0.02 M triethanolamine-acetate, pH 7.5). Protein concentrations were ca. 1.5 mg/mL and scan rates 58 deg/h.

Case. The fact that  $\Delta G_{AB}$  is only 12% as large as  $\Delta H_{AB}$  shows that most of the enthalpy change is compensated by a corresponding entropy change ( $\Delta S_{AB} = -600$  eu), consistent with the widely held view that a large conformational change occurs when regulatory and catalytic subunits are mixed to form the native protein. Being so large thermodynamically, this structure change may extend beyond the interface region and into the body of the subunits themselves.

**Yeast Phosphoglycerate Kinase (PGK).** The X-ray structure of PGK, in the absence of substrate binding, shows it to exist as well-separated amino-terminal and carboxy-terminal domains that appear to interact only through a narrow hinge region which contains a small helix (residues 185-199) and the carboxy-terminal end (residues 386-415) of the polypeptide chain (Bryant et al., 1974; Blake & Evans, 1974). A critical interaction in the hinge region appears to be the salt bridge formed between His-388 and Glu-190 (Watson et al., 1982).

Although wild-type PGK shows only a single DSC transition under most ionic conditions (Hu & Sturtevant, 1987), this transition uncouples into two overlapping transitions at pH 7.5 if only the acetate anion is present at very low concentration (Lin et al., to be published). A scan on wild-type PGK under these conditions is shown as the lower trace in Figure 3 along with the deconvolution results showing two overlapping transitions. Ligand binding experiments (discussed later) show that the low-temperature A transition is occurring in the amino-terminal domain of PGK and the B transition is in the carboxy-terminal domain.

According to the model, mutations that alter only the  $\Delta G_{AB}$  term (and not  $\Delta G^*_A$  or  $\Delta G^*_B$ ) will result in shifts in  $T_A$  but no change in  $T_B$ . We have thus far examined six mutant forms

<sup>2</sup> Although it is possible to separate catalytic and regulatory subunits in the presence of mercurials, the ATCase complex is so tightly coupled under normal conditions that it has so far proved impossible to determine the dissociation constant for  $c_6r_6 = 2c_3 + 3r_2$  by direct methods (Subramani & Schachman, 1980; Bothwell & Schachman, 1980a,b; and references cited therein). On the assumption that this dissociation is coupled to unfolding of the regulatory subunit at the concentration (ca. 10 mg/mL) of the DSC experiments, then the dissociation constant can be estimated from the  $\Delta G_{AB}$  term at 51 °C as ca.  $10^{-34}$  in molarity units. This means concentrations of native ATCase would have to be ca.  $10^{-9}$  M to observe 50% dissociation into its folded subunits and substantially lower than that at room temperature because of the large  $\Delta H_{AB}$ .

of PGK that have been modified in the hinge region. Four of these (Asp-190, Gln-190, Lys-388, and Ala-388 PGK) have changes that should alter interactions in the important Glu-190 to His-388 salt bridge in wild-type PGK while the other two (Del [413-415] and Pro-183 PGK) are modified elsewhere in the hinge area. All of these mutants exhibit nearly the same  $T_B$  as wild-type PGK, but each shows a substantially lower  $T_A$  (a shift of 5–15 °C, depending on the mutant). Single scans of three of these mutants are shown in Figure 3. Calculations from the model indicate that  $\Delta G_{AB}$  for each of the mutant forms has been significantly reduced, ranging up to a few kilocalories for those (Gln-190 and Ala-388) that no longer possess the ability to form the critical salt bridge. Even though the binding sites for substrates MgATP and 3-phosphoglycerate appear to be fully functional (i.e.,  $K_M$  values are only marginally affected) in Gln-190 and Ala-388 PGK, both mutants show considerably lower activity than the wild-type form (Mas et al., 1987, 1988). Since the enzymic mechanism is thought to involve domain movement resulting in cleft closure (Banks, 1979; Anderson et al., 1979; Pickover et al., 1979), it seems possible that the reduction of  $\Delta G_{AB}$  might make this process more difficult. This will be discussed in more detail in a later paper.

**Band 3 of Human RBC Membrane.** Perhaps the best way to apply the model and obtain an estimate of the total value of  $\Delta G_{AB}$  for multidomain proteins is by carrying out a crossover of transitions, as illustrated in Figure 2. Although this should be quite possible to do for certain proteins, there is little data in the literature indicating that it has been attempted on soluble proteins. It has been carried out in two instances for band 3, the transmembrane anion-transport protein of erythrocyte membranes. This protein has two DSC transitions: the  $B_2$  transition, which involves the cytoplasmic domain, and the C transition, which occurs in the membrane-spanning domain that includes the anion-transport channel (Brandts et al., 1978; Snow et al., 1978, 1981). The  $B_2$  transition is strongly pH dependent and changing the pH from 7.4 to 6.5 at high ionic strength (310 mOsm) moves the  $T_M$  for this transition from several degrees below the C transition to several degrees above (Brandts et al., 1978; Lysko et al., 1981) with no upward shift in the C transition resulting from the crossover.<sup>3</sup> At this latter pH, binding of the covalent anion-transport inhibitor DIDS to band 3 increases the  $T_M$  for the C transition by ca. 10 °C, which causes it to cross over the  $B_2$  transition (Appell & Low, 1982; unpublished observations from this laboratory), and again there is no detectable change in the  $T_M$  for the  $B_2$  transition. These negative results from crossover indicate little or no interaction between the transmembrane and cytoplasmic domains of band 3. These two domains of band 3 can also be separated from one another in their folded state by mild proteolysis. DSC results on the separated domains, relative to intact band 3, also suggest (Appell & Low, 1982) minimal interaction.

**Uncoupling the Thermal Transition of RNase A.** There are a number of proteins whose crystal structure shows two domains separated by a deep cleft. Some of these (RNase A, lysozyme, chymotrypsinogen, parvalbumin) have been reported to always exhibit a single cooperative transition, some (papain,

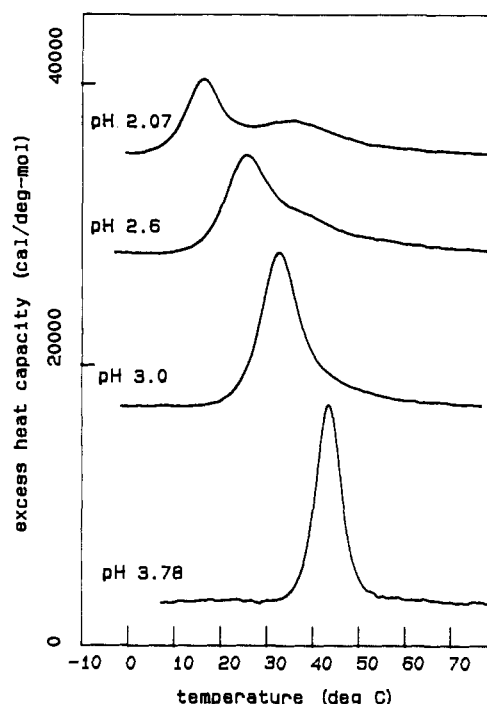


FIGURE 4: DSC scans on ribonuclease A in 50% (v/v) methanol at four different values of apparent pH. Deconvolution results (not shown) are consistent with a single two-state transition at pH 3.78 and with two overlapping two-state transitions at the three lower pH values. Second and third scans carried out on each sample indicated nearly 100% reversibility. Scan rates were 70 deg/h.

phage  $\lambda$  repressor) always show two uncoupled transitions, while still others (hexokinase, PGK) show a single coupled transition under some conditions and two transitions under other conditions. Since coupling of transitions depends on differences in intrinsic domain stability as well as on the interaction between domains, these results by themselves may provide little or no information as to which of these proteins have the strongest domain interactions.

For proteins that exhibit only one transition but nevertheless consist of two structural domains with a relatively weak seam, it is conceivable that the transition can be uncoupled either by making  $\Delta G_{AB}$  less negative, by making unequal changes in  $\Delta G^*_{A}$  or  $\Delta G^*_{B}$  such that  $T^*_{A}$  and  $T^*_{B}$  become less equal, or by carrying out both of these changes simultaneously. Since subunit interactions (and, possibly, domain interactions generally) are thought to be dominated by hydrophobic stabilization in many cases, and since different domains will frequently have different relative hydrophobic stabilization, it is possible that one or both of the above factors will move in the uncoupling direction by the addition of additives, such as aliphatic alcohols, which are thought to weaken hydrophobic bonds.

This idea has been tested (Lin and Brandts, to be published) on RNase A, long considered the standard for single-transition proteins. We have found by DSC that the RNase transition can in fact be split into two transitions by either raising the amount of methanol in the solvent at constant apparent pH or by lowering the apparent pH for RNase solutions containing 50–70% methanol. Some data obtained by the latter method using 50% methanol are shown in Figure 4. At pH 3.8 and above, the transition profile deconvolutes quite accurately as a single two-state transition with a  $\Delta H$  of about 110 kcal/mol, which is similar to the situation in pure water. As the apparent pH is lowered, the transition begins to broaden and the transition envelope can only be fit accurately by assuming two overlapping two-state transitions with approximate  $\Delta H$  values

<sup>3</sup> There is in fact a small downward shift in  $T_M$  for the C transition as the  $B_2$  transition crosses over it when the pH is lowered at 310 mOsm ionic strength (Brandts et al., 1978). At first glance, this might seem to arise from a small positive  $\Delta G_{AB}$ , but that seems not to be the case since, at lower salt (77.5 mOsm) where the  $B_2$  transition also shows a pH-dependent shift in  $T_M$  but does not cross over the C transition, the  $T_M$  for the C transition shifts downward by about the same amount as it does at higher salt.

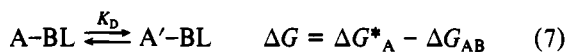
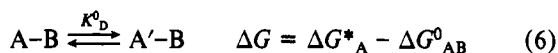
of (depending on pH) 60–80 and 50–40 kcal. Both transitions are reversible. If the methanol concentration is raised to 70% at apparent pH 3.0, all overlap between the transitions is removed since  $T_M$  for the small transition is ca. 45 °C while that for the large transition is below 0 °C (data not shown).

It is known that RNase retains activity at low temperature in solvents containing up to 50% methanol (Fink & Grey, 1978) and appears to be in a structure similar to that existing in pure water (Biringer & Fink, 1982). If this is true, there is the distinct possibility that the two DSC transitions seen in high methanol may correspond to separate unfolding of two nativelylike domains of RNase brought about by weakened domain interactions, greater disparity between intrinsic domain stabilities, or by both of these factors.

### Ligand Binding to Multidomain Proteins

In carrying out the following derivation for ligand binding to a two-domain protein, the binding domain is assumed to be the B domain and the regulatory domain the A domain and it is further assumed that the native protein always melts first to A'-B and then to A'-B' along the temperature axis. Other possibilities will be considered subsequently.

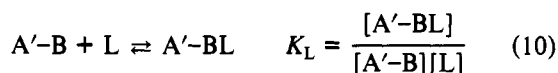
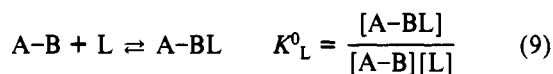
The unfolding of the A domain in the absence and presence of ligand L may be represented as



$$\Delta G_{AB} - \Delta G_{AB}^0 = RT \ln (K_D/K_D^0) \quad (8)$$

where  $\Delta G_{AB}$  and  $K_D$  are the domain interaction term and unfolding equilibrium constant in the presence of saturating L and  $\Delta G_{AB}^0$  and  $K_D^0$  are the corresponding terms in the absence of L. It will be seen later that the A domain will regulate binding to the B domain whenever  $\Delta G_{AB}$  and  $\Delta G_{AB}^0$  are different.

The corresponding binding equilibria may be represented as



where  $K_L^0$  refers to the binding of L to B when A is in the native form and  $K_L$  is the same binding constant of L to B when A is in the unfolded form. Since reaction 6 minus reaction 7 is identical with reaction 9 minus reaction 10, then it must also be true that

$$\Delta G_{AB} - \Delta G_{AB}^0 = RT \ln [K_L/K_L^0] \quad (11)$$

The experimental equilibrium constant that is monitored by DSC in the presence of ligand is

$$K_{ex} = \frac{[\text{denatured}]}{[\text{native}]} = \frac{[A'-B] + [A'-BL]}{[A-B] + [A-BL]} = K_D^0 \left( \frac{1 + K_L[L]}{1 + K_L^0[L]} \right) \quad (12)$$

$$\ln [K_{ex}/K_D^0] = \ln [(1 + K_L[L])/(1 + K_L^0[L])] \quad (13)$$

The two equilibrium constants in the ratio  $K_{ex}/K_D^0$  must be evaluated at the same temperature, and we will choose  $T_A^0$  (the transition midpoint in the absence of ligand) where  $K_D^0 = 1$ .  $K_{ex}$  is unity at  $T_A$  (the transition midpoint at concentration [L]) and may be determined at  $T_A^0$  by integration

$$\int_{T_A}^{T_A^0} d[\ln K_{ex}] = - \int_{T_A}^{T_A^0} \Delta H_{ex}/R d(1/T) \quad (14)$$

where  $\Delta H_{ex}$  is the experimental enthalpy change for the A transition in the presence of ligand [L]. If it is assumed that the transition has a temperature-independent  $\Delta C_p$ , then eq 14 integrates to

$$\ln K_{ex} = -(\Delta H_A/R)(1/T_A^0 - 1/T_A) + (\Delta C_p/R)(\ln [T_A^0/T_A] + T_A/T_A^0 - 1) \quad (15)$$

$$\ln K_{ex} = \ln [(1 + K_L[L])/(1 + K_L^0[L])] \quad (16)$$

where  $\Delta H_A$  is the experimental enthalpy change at midpoint  $T_A$ . Whenever  $\Delta C_p$  is small, a good approximation to  $K_{ex}$  at  $T_A^0$  may be obtained by using only the first term on the right-hand side of eq 15.

By measurement of the midpoint of the A transition as it changes with concentration of a ligand [L] which binds to the B domain, both  $K_L$  and  $K_L^0$  may be obtained at  $T_A^0$  by curve-fitting  $K_{ex}$  values (obtained from eq 15) to eq 16. Note that the shift in  $T_A$  with increasing [L] is a *saturable effect*, since the right-hand side of eq 16 must become constant when both  $K_L[L]$  and  $K_L^0[L]$  are much larger than unity. The change in domain interactions caused by ligand saturation,  $\Delta G_{AB} - \Delta G_{AB}^0$ , may then be evaluated at  $T_A^0$  from the limiting value of  $K_{ex}$  by using eq 11.

According to the original assumption that the A domain is always unfolded before the B domain, then the only binding constant that will be important in shifting  $T_B$  is that of L to A'-B ( $K_L$  in the above terminology), and binding to the native form is irrelevant. If it is further assumed that ligand cannot bind to the protein at all when the B domain is unfolded, then it is easy to see by analogy with eq 15 and 16 that the shift in  $T_B$  with ligand concentration is given by

$$-\ln (1 + K_L[L]) = -(\Delta H_B/R)(1/T_B^0 - 1/T_B) + (\Delta C_p/R)(\ln [T_B^0/T_B] + T_B/T_B^0 - 1) \quad (17)$$

where  $T_B^0$  is the transition midpoint in the absence of ligand. Note that the shift in  $T_B$  is a *nonsaturable effect* so that  $T_B$  will continue to shift with [L] even after all sites are saturated. This is a consequence of the assumption that the totally unfolded form cannot bind L and therefore represents an entropy of dilution effect whenever L is released upon formation of A'-B'. Equation 17 also applies directly to single-domain proteins (Becktel & Schellman, 1987) as well as to two-domain proteins when the binding domain has the lower  $T_M$ . Many proteins are known to exhibit this nonsaturable effect up to very high ligand concentrations (Edge et al., 1985; Fukada et al., 1983; Bryant et al., 1974).

For the original system under consideration, if both  $T_A$  and  $T_B$  are measured experimentally at various [L] and if the A domain is regulating the binding to B, it can be immediately recognized by a shift in  $T_A$  which may become saturated at ligand concentrations below those that cause a detectable shift in  $T_B$ . By appropriate curve fitting, the value of  $K_L$  at  $T_B^0$  may be obtained from eq 17 as well as the value of both  $K_L$  and  $K_L^0$  at  $T_A^0$  from eq 16.<sup>4</sup>

<sup>4</sup> If DSC data are accurate enough, the heat of binding of L to A'-B can be estimated from the difference in  $\Delta H$  for the B transition in the presence and absence of site saturation, which then permits a check on the internal consistency for the two estimates of  $K_L$  at different temperatures. The value of  $K_L^0$  obtained from DSC measurements should agree with estimates of the binding constant to the native protein obtained from other methods, such as equilibrium dialysis, if extrapolations are made to  $T_A^0$ .



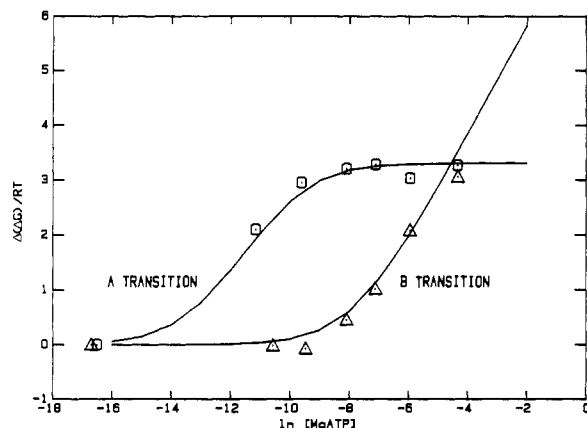


FIGURE 5: Results from  $T_M$  measurements on Ala-388 phosphoglycerate kinase in the presence of MgATP. The points represent experimental values for both the A transition (from eq 15 and 16) and B transition (from eq 17). See text for details. Scan rates were 58 deg/h.

Many two-domain proteins are known to have binding sites in the interdomain cleft. A ligand bound to one domain could then have proximal interaction with the other domain so the  $\Delta G_{AB} - \Delta G^0_{AB}$  term may, in these instances, contain contributions from direct interaction of the ligand with the regulatory domain as well as from ligand-induced changes in domain interactions, as discussed below for PGK.

#### Comparison with Experimental Results from Ligand Binding Studies

**Phosphoglycerate Kinase.** The binding site for MgATP is in the cleft on the surface of the carboxy-terminal domain of PGK, and the binding site for the other substrate, 3-phosphoglycerate, is in the cleft on the surface of the amino-terminal domain (Blake et al., 1972; Banks et al., 1979; Watson et al., 1982). Although each binding site is well separated from the neighboring domain in the free enzyme, it appears that substrate binding results in hinge motion which closes the cleft. This may happen to a greater extent when both substrates are bound than when only one is bound (Blake et al., 1972; Pickover et al., 1979).

DSC scans on the Ala-388 mutant of PGK (cf. Figure 3) have been carried out at different concentrations of MgATP from 0 to 0.013 M. At lower concentrations from 0 to 0.0003 M, increasing concentrations cause an upward shift in  $T_A$  from ca. 40 to 49 °C with no significant change in  $T_B$ . This shift in  $T_A$  saturates at low concentrations, and  $T_B$  shifts higher as [MgATP] is increased above 0.0003 M. This is the result expected if B corresponds to the binding domain (i.e., the carboxy-terminal domain of PGK) and A is the nonbinding, regulatory domain.

These data have been analyzed in terms of eq 17 (for the B transition) and eq 15 and 16 (for the A transition), and the resulting data are plotted in Figure 5. Since the abscissa is  $\ln [MgATP]$ ,<sup>5</sup> the points obtained in the absence of ligand are artificially plotted toward the extreme left in this figure. These data can be fit reasonably well, assuming a value of ca. 2500  $M^{-1}$  for  $K_L$  at 53.8 °C for the  $T_B$  data and values of 4000 and 100 000  $M^{-1}$  for  $K_L$  and  $K^0_L$ , respectively, at 39.8 °C for the  $T_A$  data.<sup>6</sup> Although this estimate of  $K^0_L$  is within the range

<sup>5</sup> Since [MgATP] refers to the free concentration of ligand, the experimental points at low [MgATP] have been shifted slightly on the abscissa in accordance with the binding constants.

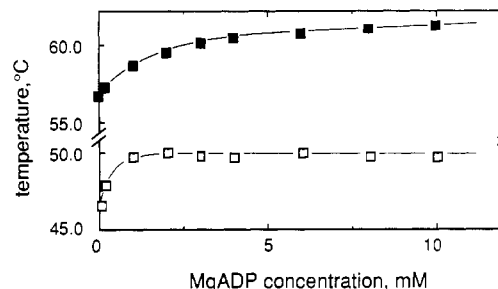


FIGURE 6: Results from  $T_M$  measurements on creatine kinase in the presence of MgADP. The points are experimental  $T_M$  values (lower curve shows  $T_A$  values and upper curve shows  $T_B$  values), while the solid lines are calculated curves from eq 17 for  $T_B$  and eq 15 and 16 for  $T_A$ .

of values reported by others for the binding of MgATP to native wild-type PGK at lower temperatures, a careful comparison cannot be made since it is known that the binding constant depends strongly on exact buffer composition (Scopes, 1978) and there is wide variation among reported literature values.

The DSC results show unequivocally that MgATP is still able to bind to the B domain after the A domain has unfolded, even though the binding constant is about 25-fold weaker. If the binding constants are converted from molar standard states to mole fraction standard states to remove entropy of dilution effects (Kauzmann, 1959), then these results show MgATP interactions of -9500 cal with native Ala-388 PGK and -7500 cal with the form having the nonbinding domain unfolded. This suggests that about 80% of the favorable interactions of MgATP with the native form are directly with the binding domain and about 20% arise from the regulatory domain (i.e., either from ligand-domain interactions or from ligand-dependent changes in domain-domain interactions). These DSC results are in accord with the accepted mechanism for PGK action, whereby substrate binding is thought to increase domain interactions thereby closing the cleft.

DSC studies of mutant PGK in the presence of the other substrate, 3-phosphoglycerate, show the pattern of behavior predicted when the binding domain is the low-temperature A domain; i.e., progressive increases of ligand concentration shift  $T_A$  higher (with no initial change in  $T_B$ ) and higher until the two transitions couple into a single two-state transition. This coupled A + B transition continues to shift higher up to the highest concentrations of ligand employed, with no crossover observed. Detailed results will be presented later (Lin et al., to be published).

**Creatine Kinase (CK).** CK is a tightly coupled homodimer (Yue et al., 1967) thought to have two active sites but perhaps only half-of-the-site activity (Degani & Degani, 1980). DSC studies (Hu and Brandts, to be published) show a two-transition pattern with a small A transition centered at 46 °C (pH 7.5) and a much larger B transition at 57 °C. Deconvolution results show a single A domain per dimer, but two identical, noninteracting B domains, consistent with a B-A-B dimer structure.

<sup>6</sup> The absolute values estimated for  $K_L$  and  $K^0_L$  from the  $T_A$  data are subject to large uncertainties but the ratio of about 25 for  $K^0_L/K_L$  is more reliable. The reason for this is that the ratio depends only on accurately locating the plateau region defined by five experimental points on the  $T_A$  curve in Figure 5 while the absolute values rely on the single experimental point on the rising portion of the curve. The two estimates for  $K_L$  of 2500  $M^{-1}$  at 53.8 °C from  $T_B$  data and 4000  $M^{-1}$  at 39.8 °C from  $T_A$  data are in approximate accord with the heat of binding of -7.5 kcal reported by Hu and Sturtevant (1987) for MgATP binding to wild-type PGK at pH 7.0 in PIPES buffer.

Addition of the substrate MgADP causes the same qualitative effects on CK previously described for MgATP interaction with PGK, i.e., a saturable shift in  $T_A$  at low [MgADP] followed by a nonsaturable shift in  $T_B$  at high [MgADP]. This identifies the A domain as the regulatory domain and the B domains as the binding domains [CK is known from other studies (Price & Hunter, 1976) to have two identical nucleotide-binding sites]. These data are plotted in Figure 6, in a slightly different form than used earlier for PGK (since too few points were obtained at low [MgADP] to be meaningfully displayed on a logarithmic scale). The solid curves are those calculated from eq 17 for the B domains and eq 15 and 16 for the A domain, using a  $K_L$  value of 4900 M<sup>-1</sup> at 56.5 °C for  $T_B$  data and  $K_L$  and  $K_L^0$  values of 4900 and 9500 M<sup>-1</sup> at 46 °C for  $T_A$  data.<sup>7</sup> This analysis is consistent with the idea that binding of MgADP to CK results in an increase in domain interactions of 800 cal (a change in  $\Delta G_{AB}$  of -400 cal per A-B interface). Even though 95% of the unitary free energy for binding MgADP to native CK results from direct interactions with the binding domain, the additional stabilization arising from the regulatory domain is still sufficient to increase the binding constant by ca. 100%.

#### SUMMARY

The fundamental assumption of the model, that both domains must be in their native state for interactions to persist, seems reasonable. Although the model is too simple to be completely accurate, it is encouraging that DSC results on complicated proteins such as ATCase, PGK, and CK in both the absence and presence of substrates are in accord with expectations. Being based on pairwise interactions, it is a straightforward matter to extend the model to proteins with more than two domains.

The unique value of the model is that its use can, under favorable circumstances, allow estimates to be made for domain interaction parameters as was shown here for ATCase (from data on isolated subunits) and for band 3 (from transition crossovers). These parameters are very difficult to evaluate by other techniques. Application of the model will be easiest for proteins where interactions are weak to moderate. For proteins with strongly interacting domains, it will be more difficult to accomplish transition crossover and/or to obtain noninteracting domains in the reference state. Even in cases where total interaction parameters cannot be determined, it may often be feasible to obtain estimates of the changes in  $\Delta G_{AB}$  that occur as the result of mutations, as was shown for PGK, or from changes in solution conditions.

It is known for many multidomain proteins (e.g., hemoglobin, ATCase) that ligand binding is stabilized not only by direct interactions at the binding domain but also by changes in the pattern of domain interactions that result from ligand binding. The studies presented here on the binding of MgATP to PGK and MgADP to CK represent the first instances where it has been possible to numerically estimate the individual contributions from each of these two factors. Results such as these can be useful not only in furthering the understanding of enzymic mechanisms for multidomain proteins but also in designing mutant proteins with altered binding properties.

Many proteins that possess a multidomain X-ray structure exhibit only a single DSC transition. This presents a problem

to the researcher interested in learning about domain structure since only composite properties of the protein can be studied and all potential information about individual domains, their binding sites, and their interactions with one another is lost. There are several ways in which transitions might be uncoupled, thereby allowing the study of individual domains. First, addition of a ligand that binds to one domain only can act to alter its relative intrinsic stability and thereby possibly uncouple transitions. Extremely large shifts in  $T_M^*$  can be accomplished (cf. eq 17) if the binding constant is large and the ligand solubility is high. Ligands that bind in the cleft between two domains are less promising uncoupling agents than those binding away from the cleft, since the former will often act to enhance the  $\Delta G_{AB}$  term (as seen in the binding of MgATP to PGK), which in itself produces a contrary effect. Second, selective mutations in the body of one domain that act to change relative intrinsic stabilities will sometimes push toward uncoupling, while mutations in the interface region that reduce the  $\Delta G_{AB}$  term (as for PGK) will always do so. Finally, changes in solvent conditions may in some instances promote uncoupling. This was seen dramatically for RNase A where the addition of methanol appears to completely uncouple its thermal transition.

#### REFERENCES

- Anderson, C. M., Zucker, F. H., & Steitz, T. A. (1979) *Science (Washington, D.C.)* **204**, 375-380.
- Appell, K. C., & Low, P. S. (1982) *Biochemistry* **21**, 2151-2157.
- Banks, R. D., Blake, C. C. F., Evans, P. R., Haser, R., Rice, D. W., Hardy, G. W., Merrett, M., & Phillips, A. W. (1979) *Nature (London)* **279**, 773-777.
- Becktel, W. J., & Schellman, J. A. (1987) *Biopolymers* **26**, 1859-1877.
- Biringer, R. G., & Fink, A. L. (1982) *J. Mol. Biol.* **160**, 87-116.
- Blake, C. C. F., & Evans, P. R. (1974) *J. Mol. Biol.* **84**, 585-601.
- Bothwell, M., & Schachman, H. K. (1980a) *J. Biol. Chem.* **255**, 1962-1970.
- Bothwell, M., & Schachman, H. K. (1980b) *J. Biol. Chem.* **255**, 1971-1977.
- Brandts, J. F., Taverna, R., Sadasivan, E., & Lysko, K. (1978) *Biochim. Biophys. Acta* **512**, 566-578.
- Bryant, T. N., Watson, H. C., & Wendall, P. L. (1974) *Nature (London)* **247**, 14-17.
- Degani, Y., & Degani, C. (1980) *Trends Biochem. Sci.* **5**, 337-341.
- Edelman, G. M. (1970) *Biochemistry* **9**, 3197-3204.
- Edge, V., Allewell, N. M., & Sturtevant, J. M. (1985) *Biochemistry* **24**, 5899-5906.
- Edge, V., Allewell, N. M., & Sturtevant, J. M. (1988) *Biochemistry* **27**, 8081-8087.
- Fink, A. L., & Grey, B. L. (1978) in *Biomolecular Structure and Function* (Agris, P. F., Sykes, B., & Loeppky, R., Eds.) pp 471-477, Academic Press, New York.
- Fukada, H., Sturtevant, J. M., & Quicho, F. A. (1983) *J. Biol. Chem.* **258**, 13193-13198.
- Hu, C. Q., & Sturtevant, J. M. (1987) *Biochemistry* **26**, 178-182.
- Janin, J. (1979) *Bull. Inst. Pasteur (Paris)* **77**, 337-373.
- Kauzmann, W. (1959) *Adv. Protein Chem.* **14**, 1-63.
- Lysko, K., Carlson, R., Taverna, R., Snow, J. W., & Brandts, J. F. (1981) *Biochemistry* **20**, 5570-5576.
- Manly, S. P., Matthews, K. S., & Sturtevant, J. M. (1985) *Biochemistry* **24**, 3842-3846.

<sup>7</sup> Although the  $T_A$  data in Figure 6 are adequate for obtaining a value of ca. 2 for the ratio  $K_L/K_L^0$ , there are an insufficient number of points at low [MgADP] to obtain the value of each. The estimates given in the text assume that the value of  $K_L$  at 46 °C is the same as that obtained at 56.5 °C from  $T_B$  data, which then allows an estimate of  $K_L^0$  to be made.



- Mas, M. T., Resplandor, Z. E., & Riggs, A. D. (1987) *Biochemistry* 26, 5369-5372.
- Mas, M. T., Bailey, J. M., & Resplandor, Z. E. (1988) *Biochemistry* 27, 1168-1172.
- Pickover, C. A., McKay, D. B., Engelman, D. M., & Steitz, T. A. (1979) *J. Biol. Chem.* 254, 11323-11329.
- Price, N. C., & Hunter, M. G. (1976) *Biochim. Biophys. Acta* 445, 364-376.
- Rao, S. T., & Rossman, M. G. (1973) *J. Mol. Biol.* 76, 241-256.
- Rossman, M. G., & Argus, P. (1981) *Annu. Rev. Biochem.* 50, 497-532.
- Scopes, R. K. (1978) *Eur. J. Biochem.* 85, 503-516.
- Snow, J. W., Brandts, J. F., & Low, P. S. (1978) *Biochim. Biophys. Acta* 512, 579-591.
- Snow, J. W., Vincentelli, J., & Brandts, J. F. (1981) *Biochim. Biophys. Acta* 642, 418-428.
- Sturtevant, J. M. (1987) *Annu. Rev. Phys. Chem.* 38, 463-488.
- Subramani, S., & Schachman, H. K. (1980) *J. Biol. Chem.* 255, 8136-8143.
- Watson, H. C., Walker, N. P. C., Shaw, P. J., Bryant, T. N., Wendell, P. L., Fothergill, L. A., Perkins, R. E., Conroy, S. C., Dobson, M. J., Thite, M. F., Kingsman, A. J., & Kingsman, S. M. (1982) *EMBO J.* 1, 1635-1640.
- Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. U.S.A.* 70, 697-701.
- Wetlaufer, D. B. (1981) *Adv. Protein Chem.* 34, 61-89.
- Yue, R. H., Palmieri, R. H., Olson, O. E., & Kuby, S. A. (1967) *Biochemistry* 6, 3204-3227.

## X-ray Studies of Aspartic Proteinase-Statine Inhibitor Complexes

J. B. Cooper, S. I. Foundling,<sup>†</sup> and T. L. Blundell\*

*Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, London WC1E 7HX, U.K.*

J. Boger

*Merck Sharp & Dohme Research Laboratories, Rahway, New Jersey 07065*

R. A. Jupp and J. Kay

*Department of Biochemistry, University College, Cardiff, Wales*

*Received January 17, 1989; Revised Manuscript Received May 4, 1989*

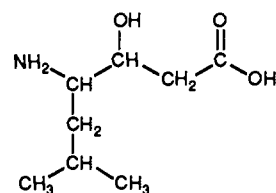
**ABSTRACT:** The conformation of a statine-containing renin inhibitor complexed with the aspartic proteinase from the fungus *Endothia parasitica* (EC 3.4.23.6) has been determined by X-ray diffraction at 2.2-Å resolution ( $R = 0.17$ ). We describe the structure of the complex at high resolution and compare this with a 3.0-Å resolution analysis of a bound inhibitor, L-364,099, containing a cyclohexylalanine analogue of statine. The inhibitors bind in extended conformations in the long active-site cleft, and the hydroxyl of the transition-state analogue, statine, interacts strongly with the catalytic aspartates via hydrogen bonds to the essential carboxyl groups. This work provides a detailed structural analysis of the role of statine in peptide inhibitors. It shows conclusively that statine should be considered a dipeptide analogue (occupying  $P_1$  to  $P_1'$ ) despite lacking the equivalent of a  $P_1'$  side chain, although other inhibitor residues (especially  $P_2$ ) may compensate by interacting at the unoccupied  $S_1'$  specificity subsite.

The aspartic proteinases are characterized by their dependence on two essential aspartic acid residues (32 and 215 in pepsin) and their susceptibility to inhibition by the microbial product pepstatin (Umezawa et al., 1970), which has an apparent  $K_i$  for pepsin of  $5 \times 10^{-11}$  M (Workman & Burkitt, 1979).

pepstatin: Ival-Val-Val-Sta-Ala-Sta

Pepstatin contains two residues of the rare amino acid statine [(4*S*,3*S*)-4-amino-3-hydroxy-6-methylheptanoic acid], the central statine being essential for potent inhibition (Rich et al., 1977, 1980; Rich & Bernatowicz, 1982).

Statine (with its leucyl-like side chain)



mimics the tetrahedral transition state or intermediate of peptide-bond hydrolysis  $[-C(OH)_2NH-]$ . In general, inhibitors having the *R* configuration at C-3 are significantly less potent than the *S* enantiomers (Rich et al., 1980; Liu et al., 1979), which indicates that the enzyme has a stereospecific binding site for the hydroxyl group. This was confirmed by X-ray crystallographic studies of complexes of fungal aspartic proteinases with pepstatin (James et al., 1982, 1985; Bott et al., 1982, 1983) which showed that the C-3 hydroxyl group is bound by a symmetrical hydrogen-bonding arrangement to the two essential carboxyl groups which probably share a single

\* To whom correspondence should be addressed.

<sup>†</sup> Present address: E. I. du Pont de Nemours & Co., Inc., Central Research and Development, Bldg 328/350A, Experimental Station, Wilmington, DE 19898.