ELSEVIER

Biophysical Chemistry 105 (2003) 545–557

Biophysical Chemistry

www.elsevier.com/locate/bpc

Uses of enthalpy–entropy compensation in protein research

Rufus Lumry*

Chemistry Department, University of Minnesota, Minneapolis, MN 55455, USA

Abstract

Cooperative systems of proteins and small molecules form most of biology but are so weakly linked that conventional mass-law formalism requiring exact stoichiometry is inapplicable. The weaknesses cannot be eliminated but using selected families of reactions useful fragmentation of those quantities is often possible. Extra-thermodynamic treatments based on linear-free-energy relationships (LFE) are developed to utilize enthalpy, entropy and volume information not otherwise reliable Linkage systems build around mesophilic proteins are well suited to enforced marriage of linear equations and scaled molecule detail because the ratio of substructure sizes on which folded stability depends is independent of total number of amino-acid residues. Conformational changes in physiological function usually no greater than 0.5 Å closely scale to linear thermodynamic changes. The formalisms for use of LFE and compensation relationships are modified to eliminate complications that have previously arisen from incorrect inclusion of the thermal parts of enthalpy and entropy changes in free energy changes. The results are used to remove current confusion about the basis of folded stability in proteins and to minimize the quantitative errors arising from classical treatments of denaturation data. The enthalpy to entropy ratio given by the slope of a compensation plot (its 'compensation temperature') is used to characterize protein construction and function so as to extract machine descriptions of protein linkage systems. In this way the 'fragile' nature of the free-energy surfaces of the myoglobin proteins and the 'strong' character of those surfaces of most other mesophiles can be deduced very simply from the Debye–Waller factors obtained in diffraction studies. The major evolutionary achievement in making proteins big is their crystallike phase behavior. That makes entropy exactly as important as enthalpy so the scalar quantities of small-molecule chemistry can be replaced by the vector quantities that appear necessary to make biology possible.

Keywords: Proteins; Evolutionary fragile and strong proteins; Temperature factors; Compensation temperature; Folded stability

1. Introduction [1]

In our first contribution to this issue [2] the usefulness of enthalpy, entropy and volume data is shown to be limited by unavoidable inclusions of contributions from the thermal and water subprocesses to which the process of interest may be linked [3]. The two-state equilibrium of water is a complication limited to highly aqueous solvents but the thermal-equilibrium problem is ubiquitous in isothermal systems. Usually these complications make only small though still significant errors in the use of the free energy changes as sources of information about the processes under study but the enthalpy, internal energy, entropy and volume changes can be too much in error to make them reliable sources of information. These weaknesses are as old as thermodynamics itself but only

*Tel.: +1-612-626-7541; fax: +1-612-625-2580.
E-mail address: lumry@tc.umn.edu (R. Lumry).

0301-4622/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0301-4622(03)00065-6

recently have begun to be appreciated. They are difficult, often impossible, to remove and as a result limit several of the most popular uses of thermodynamics as is illustrated here with applications to protein biology. A major research methodology for systems of coupled processes called linkage systems in biology is to divide the total experimental enthalpy, entropy and volume data into the contributions from each subprocess. When coupling is obligatory as in any primary-bond rearrangement, classical mass-action expressions make such separations possible but any weaker coupling is not stoichiometric and alternative methods for estimating the contributions are necessary. For such cases Wyman [4] used Maxwell's relations and the Gibbs–Duhem equation to describe linkage in terms of the concentrations of reactants and products but that methodology is awkward in attempts to compute the thermodynamic details in such systems. A formal treatment based on linear-free energy phenomena (LFE) [5,6] has advantages for such non-stoichometric problems and it is a purpose of this contribution to present the simple theory that in contrast to previous attempts [7] includes the constraint imposed by thermal equilibrium. Its applications are particularly useful for protein linkage systems and its consequences must be understood for correct interpretation of denaturation data.

Linkage systems in biology consist of a driven process, the one of central interest, and one or more driver processes. Almost any such pairing can be organized to demonstrate LFE behavior and the enthalpy–entropy relationship Leffler [8] showed to be its invariant companion. One of the earliest and most famous is the $\sigma\rho$ variety introduced by Hammett. The various forms of the driver are the otho and meta substituents of aromatic parents and the chemical reactions are the driven set. The effect measures inductive charge rearrangements in the primary-bonded systems and since the latter are not stoichiometric, the LFE methodology has made research of this kind practical. In another category the solubility of selected series of substances in a single solvent or a single solute in a variety of solvents demonstrate both LFE and linear compensation behavior particularly useful in quantifying solvent–solute interactions both chemical and physical [9]. The solvent and solute sets are selected arbitrarily using the molecular information available and good linearity confirms the selection. Linear enthalpy–entropy plots are consequence of close approximation of a thermodynamic model to systematically changing molecular properties. The former scales to the latter but the scaling is not exact so the undertaking is correctly labeled extra-thermodynamic. When the scaling is accurate within small experimental errors manipulation of enthalpy, entropy, volume and free energy as though they are exact thermodynamic quantities is often assumed and often supported by consistency among the results. Extensions to the higher temperature and pressure derivatives of H of U, S and V rarely share this limited legitimacy. Proteins processes are particularly suited to this methodology because as shown first by Pohl [10,11] and then by Privalov and coworkers [12] all mesophiles are closely related through their numbers of their residues [13] so that in their melting rates and thermodynamic stability each will respond thermodynamically to changes in many independent variables in identical or nearly identical ways. Furthermore, in enzymes both construction and function appear always to be supported by the same few devices discovered in evolution [14,15] again suggesting considerable thermodynamic similarity.

2. The equations

The LFE expression applying to series, indexed i, manifesting linear enthalpy–entropy compensation behavior is

$$\Delta G_i = \Delta G_0 + f(i)g$$

with unit free-energy increment from a coupled driven process $g = h - Ts$ and advancement factor $f(i)$ for the ith member of the series. The temperature derivatives give

$$\Delta H_i = \Delta H_0 + f(i)h - Tg\left(\frac{\partial f}{\partial T}\right)_P$$

$$\Delta S_i = \Delta S_0 + f(i)s - g\left(\frac{\partial f}{\partial T}\right)_P$$

Eliminating $f(i)$ from the two equations yields the relationship

$$\Delta H_i = \Delta H_0 - T_C\Delta S_0 + T_C\Delta S_i + h\left(\frac{\partial f}{\partial T}\right)_P\left(\frac{(T_C - T)^2}{T_C}\right)$$

in which $\left(\frac{h}{s}\right)$ *is now* labeled T_c, called the 'compensation temperature'.

Because the heat-capacity dependencies of the enthalpy and entropy changes have been ignored in this illustration, the equation as written is not appropriate for large differences between T_c and T or for series in which the f values are strongly temperature dependent.

Most of the knowledge of enzyme specificity acquired through the years has been obtained from measurements of changes in turn-over rate constant or Michaelis constant in steady-state rate studies using related series of substrates or inhibitors. The method once called the Bergmann method and still a major method often does not produce linear response of the activation or standard free-energy change computed from such parameters. When $f(i)$ is some non-linear function of i, the free-energy relationship is often detectable only through its corresponding linear enthalpy–entropy relationship so many more examples of the latter than LFE examples can be found in the literature. Linearity of the plot of enthalpy change against entropy change, the 'compensation plot' usually establishes that compensation is due to a single subprocess coupled to the measured process. If so, from a linear plot both the form of $f(i)$ and its values can be determined. However, a major exception is protein melting (vide infra). The additional experiments using temperature as the independent variable are usually worth the trouble despite the thermal complications in using the enthalpy and entropy values. Because aqueous solvents interact chemically with the hydrophobic parts of substrates, inhibiters, etc. [3], this thermodynamic version of the Bergmann method is not perfect and requires additional kinds of data for correction.

According to this LFE model its companion compensation relationship is linear in i for any series having the same parent and not otherwise. The subject is phenomenological because there is rarely a rigorous theoretical basis for that condition [16]. Nevertheless, the linearity is often as exact as the data are accurate making compensation behavior an important source of information.

As noted in the previous paragraphs, the LFE and compensation equations given above are incomplete and though generally used, they can introduce major errors in the interpretation of data. The first correction is to include the thermal-equilibrium constraint and it is the most important because it applies to data for isothermal processes of any system. The failure to include the thermal correction has made the entire subject very confusing. As shown in our first contribution to this issue [2], the confusion arises because the internal energy and entropy quantities each consist of a part that contributes to the Helmholtz of Gibbs free energy and a part that contributes the heat and heat entropy but does not contribute to the free energy. The expressions are [3]

$$U = E_0 + (p.f.)^{-1}\sum_i (E_i - E_0)e^{-E_i/\kappa T}$$

and

$$S = +\kappa \ln \sum_i e^{-(E_i - E_0)/\kappa T} + (p.f.T)^{-1}\sum_i (E_i - E_0)e^{-E_i/\kappa T}$$

with the partition function

$$p.f. = \left(\sum_i^{\infty} e^{-(E_{i0})/\kappa T}\right)$$

The last terms in U and S are the average energy fluctuation and are called heat and heat entropy, respectively. Because these cancel in A and in changes in A, they are an invariant source of linear

Table 1
Apparent compensation behavior in equilibrium melting

Temperature K	Intercept (Kcal/M)	Slope TK (T_C)	Regression coefficient
278	45.55	274.7	0.9905
298	45.4	284.5	0.9606
323	19.46	330	0.990
348	7.16	340	0.990
373	5.19	344.5	0.997
398	6.17	346.7	0.998

The table is based on Fig. 1 due to Privalov and Gill. The protein collection consists of SH3 domain, BPTI, CI2, Eglin c, Protein G, Tendamistat, Ubiquitin, Rnase T1, MetJ, cytochrome *c*, barnase, Rnase A, lysozyme (HEW), Interleukin-1β, myoglobin, T-4 lysozyme, papain, chymotrypsin. ROP and pepsinogen in the original list were not included because the data are incomplete [12]. Data for 278 K are unreliable because the two-state model fails at this and lower temperatures. Data for 398 K were obtained by extrapolation from lower temperatures to well beyond the maximum stability temperatures and have no physical significance. There were no outliers so within the errors suggested by the regression coefficient values these proteins demonstrated the same behavior. Since enthalpy and entropy were measured independently for the data Privalov and Makhatadze used in constructing this table, the statistical artifact frequently found to demonstrate spurious compensation behavior is not responsible for the accurate fitting shown by the table.

compensation behavior. As first reported by Benzinger in 1967 [17], the thermal terms do not contribute to the free energy change in any isothermal process. Only the potential energy and zero-point vibrational energies, E_0, and the degeneracy term, $+\kappa \ln \sum_i e^{-(E_i - E_0)/\kappa T}$, contribute to A so Frank suggested they should be called 'motive enthalpy' and 'motive entropy' after Carnot to distinguish work terms from the heat terms now called 'thermal' terms [18]. Designating the former with subscript m and the latter with subscript t the compensation-temperature expression for G is $T_c = \frac{h_m + h_t}{s_m + s_t}$ and the thermal-equilibrium constraint imposes the condition $h_t/s_t = T$.

The free energy change in the driven process can be computed from $\frac{h_m}{s_m}$ which is related to the mean experimental temperature T by a rearranged form of the T_c expression $\frac{h_m}{s_m} = T_c - (T - T_c)\frac{s_t}{s_m}$.

At one limit $T_c = T$ at all temperatures either because the motive terms are negligible or because the thermal and motive terms scale by the same factor. In either of these unusual cases the free-energy increment is zero. This is the basis of a 'temperature test' for this case is opposed to intermediate cases and the other extreme limit in which because there is no heat-capacity change in the driven process, T_c is independent of temperature. In that limit the free energy increment for any member of the group is equal to $\Delta\Delta G(T) = (T_c - T)\Delta\Delta S$ and the zero heat-capacity change means that for any member neither of the motive quantities changes with temperature. Zero heat-capacity change is very rare but it is found in several protein processes that determine folded stability [26] as is discussed in the following section.

3. Properties of knots

Pohl found that the activation heat capacity change for the melting rate of his mesophilic proteins to be zero independent of temperature and protein size [10]. This is the rate of denaturation from the fully-folded native protein. Most mesophilic proteins for which the pure melting rates have been determined have been found to fall on Pohl's compensation plot, which has a compensation temperature of 353 K [19]. Gregory et al. [20] also found that value for the proton-exchange rates from the sites of the small substructures responsible for folded stability. Those sites form ubiquitous substructures called 'knots' because they have to be untied before any extensive unfolding can occur [21]. That compensation temperature has been found whenever knots are rate or equilibrium determining and is useful as a diagnostic device [22]. An interesting example is the finding by Morozov and Morozov [23] that the Young's modulus goes to zero at this temperature where the mesophiles lose physical strength consistent with the fact that with a few exceptions attributable to large disulfide-bond populations all have the same denaturation rates at that temperature [24]. It has been shown that this requires that within small errors the fraction of total residues in the knots is the same for most mesophilic proteins so

their melting rates differ only in the total number of residues [13] (vide infra). Thus, for the melting-rate example the driven process is the melting rate process and the drivers are the variations in the number of residues.

Murphy et al. [25] divided the total standard free-energy change into a part independent of the heat capacity $\Delta G^o_{\mathrm{comp},j}=\Delta H^o_{\mathrm{comp},j}-T\Delta S^o_{\mathrm{comp},j}$ and a part including all the heat-capacity change. $\Delta G^o_{\mathrm{hyd},j}=\Delta C^o_{P,j}\left[(T-T^*_H)-T\ln\left[\frac{T}{T^*_S}\right]\right]$. These are pragmatically standard changes for equilibrium melting but they can be identified with the forward and back absolute rate equations

$$K=\frac{k_f}{k_b}=\frac{\frac{\kappa \mathrm{T}}{h}\mathrm{e}^{-\Delta \mathrm{G}^{\mp}_{\mathrm{f,j}}/RT}}{\frac{\kappa \mathrm{T}}{h}\mathrm{e}^{-\Delta \mathrm{G}^{\mp}_{\mathrm{b,j}}/RT}};\ \Delta G^o$$

$$=\left(\Delta G^{\pm}_f-RT\ln\frac{\kappa T}{h}\right)-\left(\Delta G^{\pm}_b-RT\ln\frac{\kappa T}{h}\right)$$

$$=\left(\Delta H^{\pm}_f-T\Delta S^{\pm}_f-RT\ln\frac{\kappa T}{h}\right)$$

$$-\left(\Delta H^{\pm}_b-T\Delta S^{\pm}_b-RT\ln\frac{\kappa T}{h}\right)$$

The first of these applies to the rate constant for melting and proves to be the same compensation relationship found by Pohl. Murphy et al. [25] also noted the compensation relationship between the standard enthalpy and entropy changes of their empirical compensation relationship, labeled comp in the following identifications:

$$\Delta H^{\mp}_{f,j}=\Delta H^o_{\mathrm{comp},j}\ \text{ and }\ \Delta S^{\mp}_{f,j}+R\ln\frac{\kappa T}{h}=\Delta S^o_{\mathrm{comp},j}$$

The identifications are quantitative and establish that both equations measure the same process, the disruption of knots. Then the work required is given by $\Delta G^o_{\mathrm{comp},j}=\Delta H^o_{\mathrm{comp},j}-T\Delta S^o_{\mathrm{comp},j}$. This unusual identification arises because the single transition state is the same as a true thermodynamic state in which the stress factors in knot and matrix just disappear. It has several consequences such as being the basis for the diffusion of catalysts for proton-exchange at knot sites without significant matrix unfolding [21].

The free-energy changes for the forward and backward rates as well as for the overall melting are given for ribonuclease A in Table 3. These are characteristic of the mesophile group. Murphy et al. also discovered that for their collection of mesophiles values for $\Delta G^o_{\mathrm{comp},j}$,$\Delta H^o_{\mathrm{comp},j}$,$\Delta S^o_{\mathrm{comp},j}$ as tabulated by Privalov and Makhatadze [12] are linearly relate to the total number of residues consistent with Pohl's finding so melting rate data can be computed from that equation. The *comp* enthalpy increment they reported is 1.52 kcal per mole of average amino acid residue and the corresponding comp entropy increment is 4.32 cal/Kmole both independent of temperature and the number of residues. As shown in a subsequent section, because the ratio of those numbers is 354 K, the compensation temperature of Pohl's melting rate plot, these numbers apply to the knot residues only. The back (refolding) rate process does not depend on total number of residues in any simple way (vide infra) and its effects are lost in the noise of compensation plot computed for the overall melting process on the incorrect assumption that melting to product is a single elementary step [26]. Nevertheless, it is well demonstrated by these construction features that mesophilic proteins are much alike regardless of size so we can conclude that a major requirement for successful evolution of mesophilic proteins has been the maintenance of a fixed ratio of knot to matrix residue numbers [13] Table 1.

These results again confirm the previous deduction that folded stability of these proteins is due to the knots. It is probably attributable to the very low electrostatic potential energy in small, tightly packed atom clusters centered on very strong peptide–peptide hydrogen bonds [3,26]. Knot strength is destroyed by expansions of clusters by approximately 0.3 Å [26]. The heat capacity of the native state is only weakly dependent on temperature and since the activation heat capacity is zero, the heat capacity of the single transition state for melting is equal to that of the native state. Furthermore, the thermal parts of entropy and entropy changes for production of the transition

Table 2
Variation of the melting parameters for ribonuclease A: Computed using the absolute rate theory expressions in text

T_m K	$\Delta G_f^{\mp}$ kcal/mole	$\Delta G_b^{\mp}$ kcal/mole	$\Delta G^o_{\text{total}}$ kcal/mole	$\Delta H_f^{\pm}$ kcal/mole	$\Delta S_f^{\pm}$ cal/Kmole	$\Delta H_b^{\pm}$ kcal/mole	$\Delta S_b^{\pm}$ cal/Kmole	$\Delta H^o_{\text{total}}$ kcal/mole	$\Delta S^o_{\text{total}}$ cal/Kmole
278	25.1	15.1	10	62.7	135	10.1	−18.2	52.5	153
298	22.4	16	6.5	62.7	135	−7.6	−79	70.3	214
323	19.1	19.1	0	62.7	135	−34.2	−164	96.8	300
348	15.7	24.2	−8.6	62.7	135	−59.7	−241	122.3	376
373	12.3	31	−18.7	62.7	135	−81.4	−301	144.1	436

The melting rate quantities are labeled *f*, the refolding quantities by *b* Equilibrium data from Makhatadze and Privalov Table 3 [12]; Rate data from Pohl [10].

state are near zero leaving only the potential energy plus zero-point vibrational energies and the degeneracy entropy term. The very small thermal terms are also consistent with the small expansion into the transition state and the small changes associated with change in hydration. The activation enthalpy and entropy changes are equal to the true thermodynamic changes in the process except for the correction for the term $(\kappa T/h)$ from Eyring's absolute rate expression which is the difference between the absolute rate expression and the empirical expression of Murphy et al. Table 2 gives data computed using the absolute rate expression. Table 3 gives results using the equations of Murphy et al since the quantities, ΔG_f^o, are corrected for Eyring's pre-exponential term and to that extent are more accurate measures of the work required to break the knots.

4. Properties of matrices

The second and much larger substructure in mesophilic proteins, 85% of the total of residues, is called a matrix because it is adjustable by mutation to carry out the physiological functions of a protein [21]. It is intrinsically unstable and maintained in its functionally important form under tension by the knots [27]. Gregory and coworkers in proton-exchange experiments established for exchange at matrix sites a compensation temperature near 470 K [28]. That identifies the expansion–contraction process of matrices oscillating from relatively high atom free volumes to glasslike rigidity [29,30]. That temperature proves to be characteristic of matrix involvement in enzymic catalysis (vide infra) and values in the range from 440 to 490 K thus far have been found to identify that matrix process although passive expansion–contraction of matrices, that is, isoergonic changes, demonstrate compensation temperatures near experimental temperatures. That matter is also discussed in the last section.

The matrix process continues to escape detection because the change in atom coordinates in contraction from liquidlike to glasslike state is smaller

Table 3
Thermodynamic estimates of knot strength in ribonuclease A: Computed using the model of Murphy et al. [25] to correct for pre-exponential terms and a more accurate measures of the true free-energy changes

T_m K	ΔG_f^o kcal/mole	ΔG_b^o kcal/mole	$\Delta G^o_{\text{total}}$ kcal/mole	$\Delta H^o_{\text{total}}$ kcal/mole	$\Delta S^o_{\text{total}}$ cal/Kmole	T_c forward K	T_c back K*
278	9.9	0	10	52.5	153	354	241
298	7.3	0.8	6.5	70.3	214	354	257
323	4.1	4.1	0	96.8	300	354	267
348	0.92	7.7	8.6	122.3	376	354	279
373	−2.9	15.8	−18.7	144.1	436	#	295

Data the same as for Tables 1 and 2. *This is the compensation temperature for the rate of refolding from melted product. (see text). #Computations for 278 and 373 K are not reliable.

than the coordinate errors in the protein X-ray-diffractions studies but it can be measured with precision of 0.1 Å using the Debye–Waller factors ('B-factors') tabulated in the PDB [14]. The B factors give the most detailed information about the distribution of atom free volumes as affected by changes in independent variables including residue exchanges and the steps in physiological function. The problem with B factors at present is their variation from laboratory to laboratory and even from protein to protein in the same laboratory [31]. The variations though less that 20% of the knot values make quantitative comparisons of a protein with potentially useful modifications unreliable. Once understood the problem can be corrected but it now presents a major barrier to a unified description of protein structure and quantitative description of the matrix expansion–contraction process. Fortunately, there are other sources of information about these 'subtle' changes. That process was first found by consolidation of information from circular dichroism, ultra-violet spectra, proton-exchange, fluorescence and catalytic parameters for chymotrypsin, all of which demonstrated linear enthalpy–entropy compensation behavior [30].

Just as the melting rate constant provides very basic information about the disruption of the knots in forming the single transition state, the compensation behavior of the refolding rate constant contains information about the formation of that state from the melted species. An important difference is that in contrast to the linear relationship between knot properties and numbers of total amino-acid residues, the points in Fig. 1 are related more through some other conformational characteristics very likely the variation in cohesive strength. This is illustrated below with the comparison of the strong protein, ribonuclease A, with the fragile one, myoglobin.

In Fig. 1 the proteins share the intercepts at zero heat-capacity change when normalized by the number of residues thus demonstrating compensation behavior related to the number of residues just as found by Pohl for the knot-dissolution rate. Murphy et al. used that point to evaluate the standard enthalpy and entropy changes for their heat-capacity-independent equation for the melting

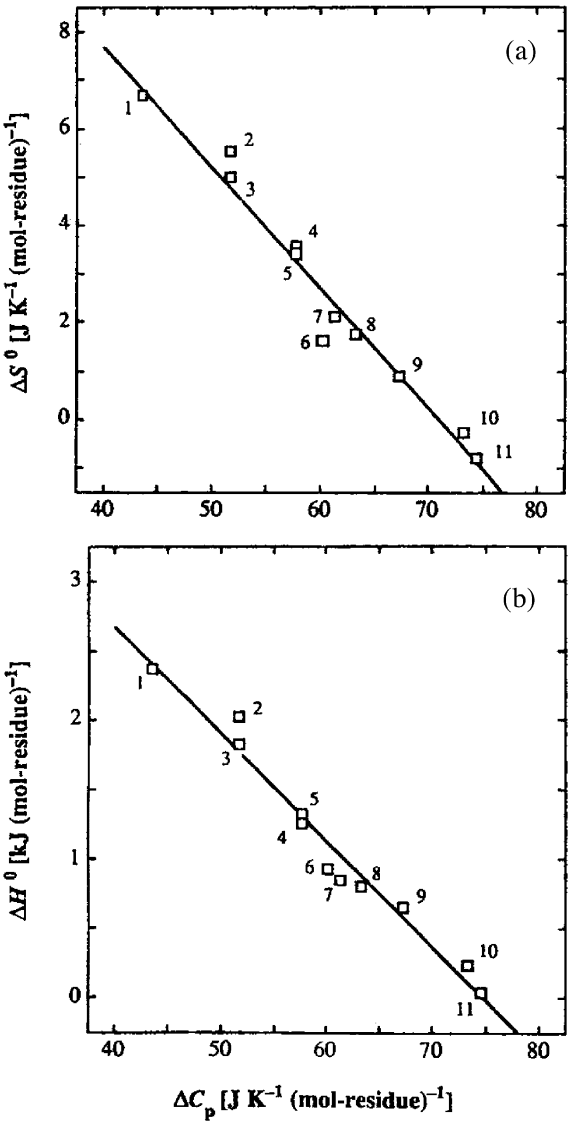

Fig. 1. Equlibrium melting data for mesophiles computed by Privalov and Gill [27] @ Linear correlations among standard changes in thermal denaturation at 298 K: (1) ribonuclease A; (2) egg-white lysozyme; (3) fragment K4 from plasminogen; (4) α-chymotrypsin; (5) β-trypsin; (6) papain; (7) *staphylococcus* nuclease; (8) carbonic anhydrase; (9) cytochrome *c*; (10) pepsinogen and (11) myoglobin.

process. The values (1.52 kcal/M and 4.3 cal/MK) determined from that common point give a ratio of 354 K for the proteins in the collection thus independent of the numbers of residues in

each. This is the same property of the proteins that lie on Pohl's compensation plot not only identifying his plot with their heat-capacity-independent equation but also adding many new points to Pohl's plot [24]. In both the heat capacity change is zero.

The number of residues in a knot of a protein may be estimated using the enthalpy increment 1.52 kcal per mole of residues by dividing the activation enthalpy for melting from Pohl's compensation plot corrected for the intercept. However, the numbers do not agree well with those estimated from the lowest B values and the number of the slowest proton-exchange sites.

4.1. Two-step mechanism of mesophile denaturation

Despite the uncertainty, the numbers computed for the back rate constant using our interpretation show correctly that the usual interpretation of protein melting data from scanning calorimetry is not correct. The data from Privalov and Gill as used by Murphy et al. are sufficient to show why this is so. Although the compensation behavior in standard melting is statistically well established (Table 1), its linearity is an artifact resulting from superposition of two different processes each described quantitatively by one the pragmatic equations of Murphy et al. Then the total melting process consists of two conformational processes one taking place in formation of the transition state from the native state and the other occurring in product formation from the transition state. In the first the tension within and between knots and matrices is released without significant change in geometry. The second involves major conformational relaxation as the compact form at the transition state expands with hydration and conformational rearrangements all contributing high motility to the compact micelle like product. The matrices change dramatically at the transition state although detectably so only in the activation entropy change in melting and in the B factors. That the product state is compact at ambient and higher temperatures [33] is a consequence of the unfavorable hydrophobic interaction Kauzmann, Edsall and Tanford made famous and is probably manifest in the same way for any large polypeptide that cannot move down the electrostatic potential-energy gradient to a native state of maximum folding, a very rare possibility made less so only by millennia of evolutionary selection.

As shown by Fig. 1, the standard enthalpy change per mole of residue rises as the standard heat-capacity increment falls, a well-known characteristic of the thermal denaturation process for these proteins and consistent with the relative stability of the matrix as estimated for ribonuclease A by the free energy change in the process shown in Table 3 as ΔG^{o}_{b}. Of this group, ribonuclease A is the most stable and myoglobin the least. Their B factors show that the former has a strong conformational free-energy surface and the latter a very fragile one. Ribonuclease A consists of two equal coils fused symmetrically into a short cylinder by disulfide bonds from coil to coil and it is probable that this use of disulfide groups provides the high stability [22]. This protein is another example of disulfide bonds used as a hinge in itself not rare but the construction to accommodate the nucleic-acid substrate is unusual despite the presence of all standard features of enzyme construction.

Since Brandts' discovery of the large heat capacity increase in protein melting [34] equilibrium melting studies and particularly those using scanning calorimetry have become a major part of protein research. His formalism for analysis of scanning calorimetry data [35] based on the two-state behavior of the protein conformation supplemented by the two-state behavior of liquid water provides accurate values of the total free energy changes during protein melting but as shown in previous sections, the enthalpy, entropy and volume changes are mixtures of motive and thermal quantities of which only the motive quantities that are reliable as sources of information about the free-energy changes. By the temperature test the standard enthalpy and entropy data for Table 1 have much larger thermal than motive parts and because of compensation, these make only minor contributions to the standard free-energy changes. Then the heat capacities would have only small between-states contributions. However, these deductions are unlikely to be correct since the

empirical equations of Murphy, Privalov and Gill correctly describe the melting process as a real two-step process though unusual in that the single one-step transition state is identical with the product of the first step. In 1993 Kim, Fuchs and Woodward [36] in a singularly important study used proton-exchange rates to detail the interaction between knots and matrices on which the separation is based. It is primarily because of the such proton-exchange data, that the two empirical equations of Murphy et al. have been identified with knot disruption and product formation following knot disruption, respectively.

As shown above, the temperature test applied to the activation quantities for the melting rate shows that knot disruption measure by that rate is entirely motive regardless of mesophile size. This is consistent with the very small value of the activation heat capacity. The contributions from the second step can be estimated from those from the melting rate and the equilibrium thermodynamics This is necessary because until very recently heat capacity data from scanning calorimeters have not been sufficiently accurate to determine the equation for the second step. Instead of using the empirical equation, one can compute the ratio of the activation enthalpy to activation entropy for the back reaction as a function of temperature. The results for ribonuclease A are given in Table 3 as $T_{c,\mathrm{back}}$. There are acceptable data for only a few temperatures and the accuracy of the estimates is not high but the ratios for the reliable temperatures are well removed from most experimental temperatures. These results suggest that the matrix disruption process is dominated by the motive changes just as is the knot-disruption process. The thermal parts are too small to explain the large heat-capacity change as a within-state contribution so it must contain a large between-state term. The apparent compensation temperatures are slightly lower than 300 K and since the motive parts must have a chemical origin, the most reasonable guess is that it is the hydrophobic interaction between newly exposed protein groups and water. Lumry and Rajender's [7] review includes many values approximately 290 K reasonably attributable to that interaction. The relaxation process responsible for the heat capacity is probably attributable to changes in populations of the two water macrostates states as water accommodates the protein groups. Above 290 K in dilute buffers the free-energy change in matrix disruption is negative rather than positive so knot formation is entirely responsible for native-state thermodynamic stability. The extent of expansion and mixing in the matrix step varies from protein to protein at any set of conditions and appears to depend on the matrix cohesive strength rather than its hydrophobic interactions. The magnitude of the free-energy contribution is closely connected to the fragile/strong characteristic of the matrix free-energy surface also observable in the standard enthalpy and entropy changes in equilibrium melting (Fig. 1).

5. Information about linkage systems

Linkage in protein systems is phenomenologically simple as a byproduct of successful protein evolution. The myoglobin-fold proteins of which the hemoglobins and myoglobins are the most studied are historical examples of such linkage systems. Beetlestone's group used ferric hemoglobin from a large variety of African animals in the first extensive study of residue exchange to show that residue changes, pH changes and ligand variations with very few exceptions demonstrate linear enthalpy–entropy compensation behavior with compensation temperatures approximately 290 K [32]. George and Hanania found 310 K in the first three steps of oxygen binding by ferrous human hemoglobin [32]. These findings were very important in showing that evolution has found a way to make entropy is as important as enthalpy in proteins [3]. With the discovery of proteins evolution had the device to make the motive part of the entropy, the generalized degeneracy, as important as the potential energy in protein construction and function. This follows from the hemoglobin compensation temperatures because they are nearly equal to the mean physiological temperature. Then $T_c=(T-T_c)\left(\frac{s_t}{s_m}\right)+\left(\frac{h_m}{s_m}\right)$ so when $(h_m/s_m)=T=T_{C,}$ the free energy change is zero. This equivalence of enthalpy and entropy is rarely found in small-molecule reactions and it is ubiquitous in

protein processes only because successful evolution has found it to be a superior basis for physiological function in enzymic catalysis as in respiratory function [14].

In the ferric hemoglobin the ligand-binding, ionization and residue changes effect changes in conformation without much change in free energy. Such processes can be called 'passive' [30]. The large substate families of the myoglobins and presumably also of the four-chain hemoglobins indicate an underlying conformational free-energy surface called 'fragile' resulting from only small inter-substate free-energy barriers. This is shown by the absence of knot B factors, the high mean value of the B factors and their small standard deviations all indicative of averaging over the fragile surface [37]. Proteins with strong knots have only a few large free-energy barriers and undoubtedly a much limited numbers of substates Their free-energy surfaces are said to be 'strong' Averaging is minimal but occurs in the expansion–contraction process of matrices within severe limits set by their tight tethering to the knots. Strong proteins are another extraordinary 'achievement' in evolution much more sophisticated than any man-made polymers because of their substructure construction and probably some electronic tricks known to spiders but not yet to man. The latter is suggested by the exceptional strengths of knots more like the extraordinary properties of spider dragline silks than mesophile matrices.

The matrix expansion–contraction process demonstrates high values for compensation temperatures, in the range from 400 to 500 K. The 470 K value found by Gregory et al. for proton exchange rates at matrix sites also characterizes the formation of the pretransition state in chymotryptic catalysis [38]. The activation enthalpy and entropy changes are positive for proton-exchange and negative for formation of the enzyme-substrate pretransition state. Reductions in B factors on substrate and inhibitor binding to chymotrypsin show such high compensation temperatures to be due to matrix contraction so proton-exchange depends on transient matrix expansion as might be expected [39]. On average, movement along this reaction coordinate is centered at the midpoint of fluctuations since both expansion and contraction processes have positive activation free energies. Transient oscillations into contracted states with lower matrix potential energy appear to be the special feature of enzymic catalysis. This is the 'entactic' behavior of Williams and Vallee [40] and the structural details revealed by enzyme B factors show that it consists of closure of the dynamically matched pair of functional domains ubiquitous in enzymes [14]. When this closure is used to compress and distort substrates into tight contact with protein functional groups in a reactive pretransient state, the device has been called a 'rack' mechanism [32] but the name 'nutcracker', recently suggested by Carloni et al. [40], is more correctly descriptive. The substrate is the nut in the protein nutcracker, probably a very accurate metaphor. Entactic behavior shows that the compression forces can be large so there is some probability that any molecule able to wander into the nutcracker can undergo a non-specific primary-bond rearrangement including perhaps free-radical formation.

The enzyme mechanism advanced by Eyring, Lumry and Spikes [41] acquired new detail with the discovery of the matrix expansion–contraction process as a transient source of potential energy to close the 'nutcracker' so to supply mechanical rate enhancement. It is a vectorial process very superior in rate enhancement and substrate specificity to the scalar enhancement provided by heat activation on which small-molecule reactions depend. Some further detail has been extracted from extensive studies of chymotryptic catalysis [38]. In these studies the compensation temperature for the initial combination of enzyme and substrate is 200 K, that for the each of the catalytic steps in the Hartley–Kilby mechanism approximately 450 K but strong inhibitors are bound at 290 K for zero free-energy change despite the large matrix contractions they produce [38]. With these numbers to relate enthalpy and entropy flux to the free-energy flux the chymotryptic linkage system begins to be understood in at least semi-quantitative terms. Thus, using the LFE method of analysis of data from a suitably selected set of experiments dictated by the independent variable to be explored the relationship of an enthalpy change to its corresponding free energy change is given by $\Delta G =$

$\alpha+\left(\frac{T_c-T}{T_c}\right)\Delta H$ in which α is a constant for the series. As already noted, this relationship does not indicate real progress because T_c contains thermal as well as motive terms and only the latter actually contribute to the free-energy change. This is the difficulty making this paper necessary since reliable uses of enthalpy, entropy and volume information depend on finding a way to isolate the motive contributions. The separation sometime can be estimated using the temperature dependence of T_c, the 'temperature test' described above. When it shows that T_c closely estimates (h_m/s_m), $g=\left(\frac{T_c-T}{T_c}\right)h_m$.

Enzyme evolution has found residue compositions that on folding produce stable knot-matrix conformations able to oscillate like a simple harmonic oscillator. In linked protein systems most of the linked subprocesses are coupled not directly to each other but through common linkage to the matrix process [16]. In contrast to the hemoglobins most other mesophilic proteins show strong free-energy surfaces and in enzymes those force coordination of conformational motions in matrices into this single degree of freedom. Thus by residue selection rather than any intrinsic feature of polypeptides the phase space available to a protein is greatly limited and each protein resembles a small crystal of a small molecule in which entropy reduction is due to crystal packing. The analogy is very close since such a crystallite no larger than a protein deviates from the artificial all-or-none mathematical model of crystallization applicable only to very large crystals by as much as the protein deviates from that model. In neither is the melting process anything like a step function. Thus, proteins are examples of Schrödinger's 'aperiodic crystals' [42] going well beyond DNA in variety if not in function.

In protein linkage systems so long as coupling is linear, as seems to be the normal situation, advancement in each subprocess can be described as an LFE expression dependent on advancement in all the other subprocesses and when subprocesses communicate with each other through protein changes, these reduce to LFE dependencies on the matrix process. The description of the linkage system is a description of the protein as a machine. Each such machine has a natural operating temperature essentially that at which the machine is in open-conformation. The coupling to each subprocess is usefully described by an enthalpy–entropy compensation process since the flow of free energy is described by the set of compensation temperatures. For illustration if the directions of advancement were the same for all component processes, those with lower compensation temperature than the operating temperature gain free energy in the operation of the machine when those higher losing free energy or vice versa. This description is detailed elsewhere and is sketched here to illustrate the usefulness of compensation temperatures. However, it has become clear that their use is limited by uncertainty in the relative values of motive and thermal terms in the compensation temperatures. In proteins motive contributions from some of the water-dependent processes are important to machine operation but the minimum of data necessary to separate them out of the total conformation process is still incomplete. The good linearity found in matrix compensation behavior can be attributed to the small geometric changes in the expansion–contraction process making the molecular scaling nearly as linear as the LFE. The protein situation will be considerably simplified if it can be shown that the constraints arising from tethering to the knots usually make motive terms large and thermal terms small. Systems in which the potentials are closely approximated by the mean-field model, characteristically show linear enthalpy-entropy compensation behaviour suggesting that protein linkage systems can be treated with some accuracy by that simple model.

6. Conclusion

The expanded state or states of proteins can be modified for study by changes in protein activity coefficients. For example, increasing amounts of glycerol as cosolvent cause chymotrypsin to contract thus slowing the proton exchange rates at all sites producing higher activation energies even at the knot sites [43]. The associated B-value changes

for this process of chymotrypsin have not been reported but the variations in B values with inhibitor binding, residue substitutions and the proton-exchange rates in crystals suggest that the dynamical behavior of matrices is reasonably well preserved in crystals. Proton-exchange rates can put studies of temporal behavior on a sound basis and nmr methods are especially suited for those determinations thorough not for determining the B factors themselves. Dielectric and acoustic relaxation methods have reported matrix conformational relaxation times slightly less than 1 ns at 298 K [44]. That such information is critical in explaining protein properties such as proton exchange rates is not generally appreciated. Contraction of matrices toward their glasslike limit is measurable with B factors in crystals but expanded states in free solution may present problems. Even if all but the largest B-factor changes are accommodated without change in crystal form, it is seldom clear from diffraction studies whether the crystal species is liquidlike or glasslike. Strong inhibitors like the diphenylcarbamyl group as an acylating group for chymotrypsin and pepstatin non-covalently bound by rhizopepsin do lock the protein into glasslike forms but less potent inhibitors rarely have predictable effects. Similar uncertainty arises in many site-directed-mutagenesis experiments in which residue exchanges even in matrices can produce dramatic changes throughout the protein [37]. Single exchanges in knots can destroy the knot or more than double its [24] strength so at present mutant methods tend to complicate the use of diffraction studies rather than simplifying them. Of the powerful tools for characterizing mesophilic proteins the B factors are the most informative even with the quantitative deficiencies apparent in the PDB tabulations. Fortunately, they apply with high precision in studies of proteins that like enzymes contain pairs of dynamically matched functional domains. Archaea proteins and immunoglobulins have not yet been compared in this way with the mesophiles.

The second most useful data are the melting rates and the corresponding activation quantities because they give fine detail of knot construction and knot properties. At this stage of protein research to provide an uncomplicated survey of protein classes and types the rates have to be measured in dilute buffers in the higher temperature region in which their rapidity requires special apparatus such as the temperature-jump apparatus used by Pohl [10]. The refolding rates can be measured at the same time or computed from the standard thermodynamic quantities for equilibrium melting. With such data the melting process can be characterized in terms of the two processes distinguished by Murphy et al. the first giving knot information and the second the relaxation into the melted product. This separation is useful because the geometric changes in the first are very small so native state and transition state have very similar conformations. Charge systems and response to solvent changes are usually best studied using the second process since it is the reverse of the assembly of the native state from the somewhat more expanded micellelike product and includes both protein matrix and its interaction with environment.

Acknowledgments

This work was supported by the Acorn Ventures Inc., Bellevue, WA 98004, USA.

References

[1] R. Lumry, The protein primer. (http://www.chem.umn.edu./groups/lumry).
[2] R. Lumry, Practical limitations on the use of thermodynamic data from isothermal processes, Biophys. Chem. 105 (2–3) (2002) 609–620.
[3] R. Lumry, Methods in enzymology, 259 Chapter 29, 1995.
[4] J. Wyman, S. Gill, Binding and linkage, University Science Books, Mill Valley, CA, 1990.
[5] J. Leffler, E. Grunwald, Rates and equilibria of organic reactions, Wiley, New York, 1963.
[6] E. Grunwald, Thermodynamics of molecular species, John Wiley & Sons, New York, 1997.
[7] R. Lumry, S. Rajender, Biopolymers 9n (1970) 1128.
[8] J. Leffler, Science 117 (1953) 340.
[9] A. Bem-Naim, Solvation thermodynamics, Plenum Press, 1967.
[10] F. Pohl, Habilitationsschrift 'Kinetics of reversible denaturation of globular proteins', Universities of Göttingen and Constanz 1969.
[11] F. Pohl, FEBS Lett. 3 (1968) 609; Eur. J. Biochem. 4 (1968) 373, 7146; FEBS Lett. 65 (1969) 239.

[12] P. Privalov, G. Makhatadze, Adv. Prot. Chem. 47 (1995) 305.
[13] R. Lumry, Biophys. Chem. 101–102 (2002) 93.
[14] R. Lumry, Protein–solvent interactions, in: R. Gregory (Ed.), Chapter 1, Dekker, New York, 1995.
[15] Ref. 1. Chap. 11.
[16] R. Lumry, R. Gregory, in: G.R. Welch (Ed.), The fluctuating enzyme, Chapter 1, Wiley Interscience, New York, 1986.
[17] T.H. Benzinger, Nature 229 (1971) 100; T.H. Benzinger, C. Hammer, Curr. Top. Cell. Regul. 18 (1981) 475; T.H. Benzinger, in: F. Heald (Ed.), Thermodynamics of life and growth, Chapter 1, Appleton, Century, Crof, New York.
[18] R. Lumry, in: A. Braibanti (Ed.), Bioenergetics and thermodynamics: model systems, Reidel, Dondrecht, Holland, 1980, p. 405.
[19] Ref. 1, Chap. 2.
[20] R. Gregory, Biopolymers 22 (1983) 895.
[21] R. Gregory, R. Lumry, Biopolymers 24 (1985) 301.
[22] Ref. 1., Chaps. 3 and 13.
[23] V. Morozov, T. Morozov, J. Biomole. Struct. Dyn. 11 (1993) 459.
[24] R. Lumry, J. Biophys. Chem. 101–102 (2002) 81.
[25] K. Murphy, P. Privalov, S. Gill, Science 247 (1990) 559.
[26] Ref. 1, Chap. 7.
[27] Ref.1, Chap. 8.
[28] R. Gregory, Biopolymers 22 (1983) 895.
[29] R. Gregory, in: R. Gregory (Ed.), Protein–solvent interactions, Dekker, New York, 1995, p. 191.
[30] R. Gregory, in: D. Reid (Ed.), The properties of water in foods—ISOPOW 6, Blackie, Academic and Professional, London, 1998, pp. 57–100, (http://kente-du.chemistry/moleimage).
[31] R. Lumry, R. Biltonen, in: S. Timasheff, G. Fasman (Eds.), Structure and stability of biological macromolecules, Dekker, New York, 1969.
[32] R. Corbett, R. Roche, Biochemistry 23 (1984) 1888.
[33] J. Brandts, J. Am. Chem. Soc. 86 (1964) 4291–4302.
[34] J. Brandts, in: S. Timasheff, G. Fasman (Eds.), Structure and stability of biological macromolecules, Chapter 3, Dekker, 1969.
[35] K.-S. Kim, J. Fuchs, C. Woodwad, Biochemistry 32 (1993) 9600.
[36] Ref. 1, Chap. 10.
[37] R. Lumry, in: S. Kuby (Ed.), Study of enzymes II, CRC Press, Boca Raton, Florida, 1991, pp. 36–45 and also ref. 32, pp. 154–170.
[38] V. Dorovska-Taran, R. Momicheva, N. Gulubova, K. Martinek, Biochim. Biophys. Acta 702 (1982) 37.
[39] R. Williams, B. Vallee, Proc. Natl. Acad. Sci. (U.S.A.) 59 (1968) 498.
[40] S. Plana, P. Carloni, M. Parrinello, J. Mol. Biol. 319 (2002) 567.
[41] H. Eyring, R. Lumry, J. Spikes, in: W. MeElroy, B. Glass (Eds.), Mechanisms of enzyme action, Johns Hopkins University Press, Baltimore, 1954, p. 123, R. Lumry, H. Eyring, J. Phys. Chem 58 (1954) 110.
[42] E. Schrödinger, What is life?, Cambridge University Press, Cambridge, 1947.
[43] R. Gregory, Biopolymers 27 (1988) 1699.
[44] N. Mura, Y. Hayashi, S. Mashimo, Biopolymers 39 (1996) 183.