This version is provisional because it contains several long, unmodified abstracts from two papers to appear in September 2003 in the Biophysical Journal and thus copyrighted by Elseverier Publishers. (7-15-03)

Chapter 22. Theory of non-stoichiometric linkage systems.

When two or more processes are coupled so that advancement in one produces advancement in the others, the collection is known in biology as a linkage system. Biology consists of such systems made famous by Jeffreys Wyman and in chemistry primarily by the activity formalism advanced by G. N. Lewis. The novel feature of linkage systems is the absence of the obligatory coupling that produces exact stoichiometry and the exact mass-action expressions that make so much of chemistry mathematically simple. The most common source of obligatory coupling is redistribution of atoms since changes in primary bonds involve very large energies. A classical example of linkage in biology is the interaction of the four subunits of mammalian hemoglobins which produces the physiologically useful relationships between oxygen pressure and the amount of oxygen actually bound by hemoglobin. Free energy released in one oxygen binding process on one subunit is redistributed over the remaining three subunits to increase or decrease the affinity of the other subunits for oxygen. The stoichiometry changes continuously Wyman and Wyman and Gill developed the activity relationships in terms of advancement at the ligation sites and any linkage to pH, solvent compositions, etc. These yield quantitative description of the free-energy flow that describes a linkage system as a machine doing work. That description provides little information about the construction of the machine and it is common to turn to enthalpy, internal energy, entropy and volume information for such detail although it is awkward to extract that information from the concentration relationships. Those quantities extracted from temperature and pressure dependencies or by experimental calorimetry are

ambiguous especially for polymers such as proteins. That state of affairs was not discovered until 1967 and is not yet in textbooks. The ambiguity arises from the fact that a system working at constant temperature cannot convert heat to work. The free-energy changes involve interconversions within a system or between systems in which the enthalpy and entropy changes are not zero but can make no contribution to free-energy changes so long as the temperature and pressure are held constant. The theory due to Benzinger has been given in several places but a recent discussion to be published in the Biophysical Journal (September 2003) is adequate not only not only for the constraints arising from constant temperature and by analogy constant pressure but also for coupling to the two-states of water, another source of ambiguity in the relationship of free-energy changes to enthalpy etc. That water does have two and only two macrostates long the subject of intense discussion was established most impressively by Walrafen and coworkers in 1983 but the errors due to that linkage are rarely even acknowledged. A prepublication version of our discussion of this odd omission in the history of thermodynamics is given as PDF file on the Protein Primer server as Benzinger.pdf.

For most non-obligatory linkage systems the free-energy changes cannot be corrected for heat enthalpy, entropy and volume by any known experimental methods so other methods are required to utilized those quantities of which the most useful is the extrathermodynamic pair called “linear free energy” and “enthalpy-entropy compensation behavior” These come in pairs as first shown by Leffler and they have become useful because of Benzinger discovery about th enthalpy, entropy and volume relationships to free energy. The useful is limited but the slope of linear enthalpy-entropy compensation plots gives the ratio of enthalpy change to entropy change from which unique information about conformational processes of proteins for example can be extracted. That is illustrated in the following.

It is the linearity of the enthalpy-entropy plots that makes them useful. And mass-action bases yield apparent linearity only over a very small range of change in the independent variable. That was examined by Biltonen, Eftink and Anusiem to show that the linearity is not exact. The apparent linearity commonly demonstrated is then an extra-thermodynamic consequence resulting from small errors in the data that hide the fact that the molecular changes along the series do not scale exactly with the thermodynamic changes, enthalpy, entropy or free energy. Consider the later changes in transferring small linear amphiphiles such as the linear alkanols from a passive solvent or a gas phase into water. So long as the aqueous solutions are very dilute and the alkanols not too long the changes in mass and number of hydrogen atoms are linear along the series to a high level of accuracy so they scale within those errors with the thermodynamic changes. Methanol is off the line as are any branched-chain alkanols added to the collection. Scaling begins to fail in H, S and G so that their derivatives such as heat-capacity and compressibility obtained as T or P derivatives become unreliable in the sense that the linearity becomes inexact.

The linear alkanols give linear plots within errors because the only changes along the series are in mass and number of hydrogen atoms. The noble gases on the other hand do not give linear plots because both size and polarizability change either by itself would give good linearity but together produce curvature in the compensation plots. Such non-linear scaling does not make compensation analysis useless but because better data are required and multiple sources of scaling more difficult to distinguish, non-linear compensation behavior has not been popular. Linear compensation behavior is the major too for quantitative analysis of protein data. The major source appears to be the large size of the protein and the greatly constrained conformational process due to evolutionary selection rather than any intrinsic properties of unstructured polypeptides. The small and large molecule processes providing the

physiological properties are not directly coupled to each other but rather through individual attachments to a central protein. The matrix expansion-contraction process of that protein has two-state character along a single conformational coordinate with variable but linear displacement between the extreme expanded and contracted termini. The thermodynamic changes even those of the free energy can be quite large without detectable deviations from linearity. In this way evolution has fashioned almost thermodynamic behavior from an extra-thermodynamic basis. Ligand binding can also be linear following the model of the alkanols series discussed above. Changes in protonation not only perturb the matrix of the central protein but because experimental systems are usually buffered near the acid or base constants the free-energy changes are near zero again favoring linear behavior in acid or base concenration or pH depending on the involvement of ionization with matrix or with substrate binding.

Linear enthalpy-entropy compensation behavior within suitably small experimental errors determined by the quantitative errors than can be tolerated can be treated by simple theory we now illustrate to establish the physical and chemical significance of the compensation behavior of the melting rate constant first observed by Pohl and the first appearance of the unique temperature of 354K for mesophilic proteins. That and related topics are discussed in two paper in the Journal of Biophysical Chemistry for December 2002 and are now copyrighted by Elsevier and so not available on the Protein Primer web server. Two newer manuscripts covering the topics of this chapter in greater depth but also copyrighted by Elsevier will appear in the September 2003 issue of that journal.

Extra-thermodynamic phenomena in protein chemistry.

There are several sources of linear-free-energy and enthalpy-entropy compensation behavior is protein systems. The two state equilibrium of water is

a complication limited to aqueous solvents; the thermal equilibrium problem is ubiquitous in isothermal systems. These complications make minor errors in the use of free energy changes as sources of information about the solute process under study but the T and P derivatives of the free energy often have large quantitative uncertainty... The major need in systems of coupled processes is to divide the total experimental enthalpy, entropy and volume data into the contributions from each subprocess. When coupling is very strong as in any primary-bond rearrangement classical mass-action expression make such separations possible but any weaker coupling is not stoichiometric and alternative methods for estimating the contributions are necessary. For the latter cases Wyman used Maxwell's equations and the Gibbs-Duhem equation to describe linkage in terms of the concentrations of reactants and products but that methodology is less powerful in estimating thermodynamic details of linkage. A formal treatment based on linear-free energy phenomena has advantages for such non-stoichometric problems and it is a purpose of the this contribution to present a theory that in contrast to previous attempts includes the constraint imposed by thermal equilibrium Its applications are particularly useful for protein linkage systems as is shown below for several of the most common processes of proteins.

Linkage systems 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 showed to be an 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 measure inductive charge rearrangements in the primary-bonded systems and the LFE methodology have made research of this kind possible. In another well-known 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. The solvent and solute sets are selected arbitrarily using the molecular information available and good linearity confirms the selection. A linear enthalpy-entropy compensation phenomenon is a consequence of close approximation of a thermodynamic model on true molecular properties. The former scales to the latter but the scaling if not rigorous 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 assumed and often supported by consistency among the results. Further extensions to the higher temperature and pressure derivatives of H of U, S and V rarely share this limited legitimacy

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 driver process g=h-Ts and advancement factor f(i) for the i th 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

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

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

Because the heat capacity dependencies of the enthalpy and entropy changes have been ignored, the equation as written is not appropriate for large differences between Tc and T nor for series in which the f values are strongly temperature dependent.

Because f(i) is often some non-linear function of i, the LFE relationship is often detectable only through its corresponding linear enthalpy-entropy relationship so there are many more examples of the latter than have been detected.. Since the linearity of the latter establishes that compensation is due to a single subprocess coupled to the measured process, the form of f(i)) can be determined from the compensation plot. According to this LFE model its companion compensation relationship is linear in i for any series having the same ΔH_0 and ΔS_0 and not otherwise. The subject is phenomenological because there is rarely a rigorous theoretical basis for that condition. Nevertheless the linearity is often as exact as the data are accurate.

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. As shown in our second paper in this issue 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 to the heat and heat entropy but does not contribute to the free energy. The expressions are;

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

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

The last term in each is the average energy fluctuation and is called heat and heat entropy, respectively. Because these cancel in A and in changes in A, they are an invariant source of linear compensation behavior. The potential energy and zero-point vibrational energies, E0, and the degeneracy term, $\kappa \ln \sum_i e^{-(E_i - E_0)/\kappa T}$, in the entropy contribute to A and are called "motive" quantities in contrast to the two heat terms now to be called "thermal" terms. Designating the former with subscript m and the latter with subscript t the compensation-temperature expression for G is

$T_c = \frac{h_m + h_t}{s_m + s_t}$. In one limit $T_c = \frac{h_m}{s_m}$ is rare because it requires that the heat capacity be zero ($\left(\frac{\partial s_m}{\partial T} \right)_V = \frac{s_t}{T} = 0$) so Tc is independent of temperature.

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. Although that case I
s rare a number of other proteins have been found to fall on the same compensation plot and statistical treatment gives a compensation temperature of 353K with an error no greater than ten degrees. Gregory et al also found that value for the proton-exchange rates from the sites forming the knots, the small substructures responsible for folded stability and it has been found whenever

knots are rate or equilibrium determining as found in the studies of the Young's modulus by Morozov and Morozov. With a few exceptions attributable to large disulfide-bond populations all mesophiles become unstable and 353K and it has been shown that this requires that the fraction of total residues in the knots is the same within small errors so that with respect to melting rates mesophiles differ only in the total number of residues. The driven process is the melting rate process and the driver is the number of residues. By extending the treatment of thermodynamic data for equilibrium melting of mesophiles given by Murphy et al it is found that those data are also normalized by normalizing to individual proteins to a fixed number of residues.

These results confirm the fact that folded stability is due to the knots and a consequence of the very low electrostatic potential energy of small residue clusters centered on very strong peptide-peptide hydrogen bonds. Knot strength is destroyed by expansions of only a few tenths of angstroms. Since the thermal terms are both zero, the activation free energy for melting is entirely potential energy and degeneracy entropy. This kind of substructure is called a knot because it has to be "untied" before unfolding can occur.

Matrices

The second and much larger substructure in mesophilic proteins, roughly 85% of the total, is called a matrix because it is highly adjustable to carry out the physiological functions of a protein. It is intrinsically unstable and maintained only by the knots. In thermal melting it loses structure and becomes mobile even with only small expansion. It demonstrates linear enthalpy-entropy compensation behavior and was first detected by Gregory and coworkers in proton-exchange experiments that established the compensation temperature in a wide range from □ 50 degrees about 450K. That value indicates a very different composition of motive and thermal parts, almost the opposite of that seen in the

knots. This has been shown very clearly by calculating the activation quantities for the refolding rate from denatured product to the transition state in the overall two-step folding or melting process. The motive parts are very small relative to the very large negative values of the activation enthalpy and entropy for the refolding rate from the normal melted state, a small motile disordered micelle in dilute buffers. Since the thermal terms cancel in the free-energy change, neither the entropy nor entropy change has much bearing on folded stability. The heat capacity change is very large because the thermal energy change is so large but like the latter is irrelevant in explaining stable folding...

Thus for refolding and unfolding beyond the transition state the difference between motive and thermal contributions is crucial for understanding. The experimental way to test is to use the "temperature test" because the compensation temperature changes with mean experimental temperature and the motive parts change much less so. The data on equilibrium thermal denaturation of mesophilic proteins compiled by Privalov and Makhatadze yield compensation temperatures between 285K and 350K close to the experimental temperature and even closer for the activation enthalpy and entropy plots for the refolding process. At 285K and below the two-state model for denaturation does not apply and at 360K and above the data obtainable for mesophiles only by extrapolation exceed the stability limit of 354K and are irrelevant. Compensation temperatures and their temperature dependencies are major tools in analyzing linkage systems as is shown below.

Information from compensation temperatures

The physiological functions supported by many and probably most proteins depend on matrix expansion-contraction much like the dependence of small molecule reactions on heat but in a far more versatile and powerful way. The atom B factors are not the only way to follow that process but they measure

free volume and the process is no more than change in free volume, they are the most informative source. Circular dichroism, fluorescence, phosphorescence and other measures of conformation characteristics are also useful but in general have not yet been calibrated using B factors. Compensation temperatures more one step deeper into the machinery because they characterize the matrix process in thermodynamic terms so as to report the degree of advancement of that process but also they are a highly quantitative measure of effects of change in independent variables. Site directed mutagenesis gives employment to many but new information to few until such time as it is systematically correlated with B-factor changes as they can be using existing x-ray-diffraction data.

Linkage in protein systems is phenomenologically simple because protein evolution has been so successful. The myoglobin-fold proteins of which hemoglobin and myoglobin are among the most studied examples of compensation in linkage systems. However these proteins differ from most well-known mesophiles in having small free-energy barriers among most of their conformers. In hemoglobin the atom B factors have a high mean and a small standard deviation as is consistent with averaging by easy transition from one of Frauenfelder's many substates to another. Free-energy barriers among these states must be low a condition called a "fragile free-energy surface" in contrast to the "strong surfaces" of proteins with strong knots. Although the distinction between knot and matrix is probably maintained in many of those substates as shown by the persistence of the B-factor palindrome, most detail is washed out. The protein is soft in that sense so the small adjustment of the many helices to changes in ligation at the heme group are expected to have minor influence on conformation. Beetlestone's group used ferric blood of many African animals in the first study of residue exchange to show that residue changes, pH changes and ligand differences with very few exceptions demonstrated linear enthalpy-entropy compensation behavior with compensation temperatures around 290K.

George and coworkers found similar behavior in first three steps of oxygen binding by ferrous human hemoglobin. These findings were very important in showing that entropy is as important as enthalpy in proteins. More correctly evolution has made the degeneracy part of the entropy as important as the potential energy in protein construction and function. This follows immediately from the fact that the compensation temperature is nearly equal to the experimental temperature since that means they reactions are nearly isoergonic. All contributing pairs of the entropy and enthalpy contributions are very close to equilibrium including the motive parts. This equivalence is not found in small-molecule reactions and occurs in proteins only because of very successful evolution.

The compensation temperature is equal to the ratio of motive enthalpy change to motive entropy change because of the cancellation of thermal and water contributions and this is true at any temperature equal to the compensation temperature. The situation is more complicated when that is not true but the compensation temperature is still the best indicator of the thermodynamic character of the process. It has been called the thermodynamic phase factor since melting for example has a ratio greater than unity value up to the transition state and a less-than unity from that state to product. In hemoglobins the various experiments of Beetlestone and others indicate passive bindng, that is, no major free-energy change. The matrix process of enzymes on the other hand demonstrates high values for compensation temperatures and thus major changes in free energy. The open, high potential-energy, states are strongly favored on average but transient oscillations into states of lower potential energy occur and are apparently critical to catalytic function. They are the closing of the nutcracker that by raising the potential energy of the pretransition state for primary-bond rearrangement makes rapid passage into the transition state to products possible...

What is especially important with these proteins is that the temperature test shows the compensation temperature to have must less dependence of experimental temperature than expected if the thermal terms dominated the motive terms in the standard enthalpy and entropy changes in ligand binding which were measured. This means that the ratio of motive enthalpy change to motive entropy change is close to 290K so that the standard free-energy changes are close to zero at that temperature. The system is then similar to that between simple solutes and water in. When the compensation temperature is equal to the mean experimental temperature, the ratio of motive enthalpy change to the entropy contribution, T times the motive entropy change, is unity. In most small-molecule processes the latter is usually small corresponding to a high compensation temperature and not of much use in managing free energy. In protein systems the equivalence adds a second dimension to free-energy. making possible free-energy adjustments vectorially rather than the scalar adjustments that dominate small-molecule reactions. That is a major explanation for the large size of protein molecules by detailed selection of atom free volumes in enzymes nature bypasses thermal activation with mechanical activation of rate processes the discovery that has made biology possible.

The compensation temperature as the ratio of enthalpy change to entropy change is the index of free-energy change. As such it has been called the "thermodynamic phase factor". The 354K values for mesophile knots describes thermodynamically the proton exchange rates for matrix site, the thermodynamic stability of folded forms and the Youngs modulus of native forms. In another place it is shown than the constancy of this compensation temperature requires that all mesophilic proteins have the same ratio of the number of knot residues to matrix residues and all melting rates are zero at 354K. BPTI and ribonuclease A have somewhat higher maximum at 370K apparently due to their large numbers of disulfide bonds. What is most remarkable is that when the total

number of residues is normalized to a single reference for these proteins, the activation quantities for melting are also normalized.

There is no thermal complication in the melting rate process because the increase in potential energy and degeneracy entropy takes place with knot expansions of only a few tenths of angstroms. The thermal complications arise from the subsequent process from transition state to melted product in which major changes in conformational freedom and heat occur. Those depend on the state of the matrix in the normal species since the major functions of matrices at least in mesophilic enzymes involves the matrix expansion-contraction process. Evolution has selected residue compositions that on folding are mutually constrained to behavior like a simple harmonic oscillator. In linked protein systems the linked subprocess are linked not directly to each other but through common linkage to this matrix process. In contrast the hemoglobin in which linkage are weak and inexact enzymes and most other proteins in this class have strong free-energy surfaces that force coordination of conformational motions in matrices to this single degree of freedom. By residue selection rather than any intrinsic feature of polypeptides the phase space available to a protein is greatly limited. 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 by as much as the protein deviates from that model. Thus it is the freeze-thaw (contraction-expansion) process of matrices with midpoint near 290K that directly makes biology possible.

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 but so long as 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 at rest and the coupling to each subprocess is most usefully described an enthalpy-entropy compensation process since the flow of free energy is described by the set of compensation temperatures. Those lower than operating temperature gain free energy in the operation of the machine, as a catalyst for example, and those higher losing free energy. This description is detailed elsewhere and is sketched here to illustrate the importance of compensation temperatures. Detail is not necessary in fact sophistic at this time because the expansion-contraction process itself can be related quantitatively not only in the advancement variable of the subprocesses but also to measure the consequences of substitutions in amino-acid-residues. This happy state of affairs is a consequence of the B factors as measures of the protein process. B factors for this purpose are not yet perfect since the way they scale from one to another diffraction experiment is quite variable for reasons that have not yet been explained. The B values have a sharply fixed mean with very small standard deviation in any one diffraction experiment so that the cusps of the knot palindromes match in atom positions computed from atom B values within 0.1 angstrom or better in the highest resolution modern x-ray-diffraction studies. The uncertainty in the coordinates even at the highest resolution is much larger although analysis by Fourier difference methods such as used by Moffat and coworkers provides considerable improvement. However, the accuracy is well below that required for molecular-dynamic treatments and generally no improvement over that from B factors. Bahar et al have devised algorithms to estimate local packing from coordinates that improve the use of coordinates and their errors.

Computer modeling of proteins at present depends on coordinate errors large with respect to the distance dependence of the potential energy so that the

results do not provide a reliable bridge from diffraction data to quantitative values of the thermodynamic characteristics. Until the experimental variability in the B factors can be removed, we are dependent on the empirical values of the free energy data and of course also enthalpy, entropy and volume data without an accurate bridge to the B factors and, of course, with the errors resulting from constant temperature, constant pressure and the two macrostates of waters. This is why the application of compensation theory is important despite is uncertainties and extra-thermodynamic weaknesses.

The variable responsible for the compensation behavior generated by the activation quantities for the melting rate of mesophilic proteins is the number of amino-acid residues. As already discussed its characteristics have led to the solution of the "protein folding problem" for these proteins revealing remarkable convergent evolution. The refolding rate constant has complications due to linkage to the water two-state process and to the thermal equilibrium but the contributions cancel to produce small activation free energy changes. In this case the cancellation itself is the source of the most important chemical information.

The B factors reveal the residues and atoms forming the knots and these are involved in physiological processes in a secondary way. Information about the support of those processes is primarily encoded in the matrix B factors and other observables giving B-factor information. Like any machine a protein machine is described in terms of the thermodynamics of its operation. It is the advancement in those quantities with conformational changes that measures the bridge between conformational and chemical changes in protein linkage systems and thermodynamic change. Here the compensation temperature provides essential information