Chapter 13 Similarities and differences among proteins

In dilute buffers the activation heat capacity for melting has been found to be near zero. Even in relatively low concentrations of denaturants this simple behavior is lost as illustrated by the studies of Chen and Schellman. The studies of chymotrypsinogen by Eisenberg and Schwert and soybean trypsin inhibitor by Kunitz were the first to demonstrate the zero activation heat-capacity to be followed twenty years later by Pohl's measurements for ribonuclease A and several members of the trypsin and trypsinogen families in free and acyl forms. Sugihara and Segawa added HEW lysozyme to the list. Hopkins and coworkers measured the rates for several chymotrypsins and some of their derivatives using fluorescence changes following rapid injection of aqueous solutions into 8M urea at pH 7.2. This simple method is limited because the urea rapidly destroys the B hydration shell before the transition state is reached but that is one of its virtues. With few exceptions the melting rate measures the rate of knot disruption so the activation enthalpy and entropy changes fall on Pohl's compensation plot with fixed 345K slope. Most proteins so far studied for denaturation rates in the presence only of water and dilute buffers have been found to fall on Pohl's compensation line demonstrating that they all have the same melting rate under these conditions. This simplicity is hidden in the standard free-energy change in melting by secondary effects of charge and solvent composition. Thus changes in the native state produced by temperature , pressure, residue number and exchanges of residues, ionization, solvent composition, matrix contraction, inhibitor binding and such accessible intermediates in catalysis as the acyl enzyme species formed in tryptic hydrolysis can be measured in an unambiguous way. Furthermore only the rates themselves need be measured because the activation enthalpy and entropy changes can be computed using Pohl's compensation relationship. The melting rate is the single

most informative measurement of native state properties available but many of the few experiments so far reported are ambiguous because urea or some other structure-breaking cosolvent was used to move the rates into a more easily measured time range. The urea complication can be removed by further exploitation of the method of Hopkins et al. to calibrate the co solvent effect.

Biltonen first observed that since the melting equilibrium data were well fit by the two-state model that the large standard heat-capacity increase occurs on the product side of the transition state. Murphy, Privalov and Gill assuming that the heat-capacity is due to exposure of melted protein to water divided the total standard free-energy change into a part independent of the heat-capacity $\Delta G^o_{comp,j} = \Delta H^o_{comp,j} - T\Delta S^o_{comp,j}$ and a part including all the heat-capacity change. $\Delta G^o_{hyd,j} = \Delta C^o_{P,j}\left[\left(T - T^*_H\right) - T\ln\left[\frac{T}{T^*_S}\right]\right]$. These equations were applied to data for 13 mesophilic protein tabulated by Privalov and Makhatadaze. The parameters in the second equation were determined empirically and are equal to 385K within small errors and *j* is the series index. Privalov and coworkers made several important deductions from this application that deserve more attention. The first is that the second equation is always negative for T less than 385K so that formation of the melting product is favored. Most previous conventional wisdom has been that hydrophobic interaction of the product with water is unfavorable and the major source of thermodynamic stability of the native species. Privalov correctly reasoned from the fact that the standard enthalpy change in the first equation is always positive that it is responsible. These workers incorrectely assumed that the product was an extensively unfolded into direct contact with bulk water but that does no contravene Privalov's deduction.

It is profitable to explore the meaning of the first equation since in doing so their physical significance can be established.

So long at the two-state model is an accurate fit to the melting equilibria the equilibrium constant can be expanded into Absolute rate theory expressions for the forward and backward rate constants.

$$K = \frac{k_f}{k_b} = \frac{\frac{\kappa T}{h} e^{-\frac{\Delta G^{\mp}_{f,j}}{RT}}}{\frac{\kappa T}{h} e^{-\frac{\Delta G^{\mp}_{b,j}}{RT}}}; \quad \Delta G^o = (\Delta G^{\pm}_f - RT \ln \frac{\kappa T}{h}) - (\Delta G^{\pm}_b - RT \ln \frac{\kappa T}{h})$$

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

If these replace the empirical separation of Murphy, in the rate-theory form above the first term on the right is the conventional Eyring rate theory expression for the forward rate constant and can be identified with their heat-capacity- independent equation by comparison of the two:

$\Delta H^{\mp}_{f,j} = \Delta H^o_{comp,j}$ and $\Delta S^{\mp}_{f,j} + T \ln \frac{\kappa T}{h} = \Delta S^o_{comp,j}$. However the physical validity of the identifications become useful only because Pohl's direct measurements of the melting rate are identical with the melting-rate expression. Thus in both formulations the melting rate data give linear compensation behavior with the characteristic knot compensation temperature of 354K. The extra part of the entropy in the rate theory equation is 53 cal/MK and appears in Pohl's compensation plot as the intercept of enthalpy on the entropy axis in Pohl's plot (16kcal/M at 298K). The compensation temperature has already been identified as characteristic of knots and its appearance in both the plot of Murphy et al where it is compensation plots establishes that the rate-theory version of the equation of Murphy et al derived above is identical with that found by Pohl for the melting rates. The latter is conventional rate theory so this agreement validates the computation of the rate theory parameters from the thermodynamic changes in melting equilibria, an unusual but very useful

procedure making it possible to obtain melting and folding rate data from standard thermodynamic changes at least for mesophilic proteins.

In this way we find a clean separation between knot stability measured by the melting rate constant and the several separate factors determining the rate from the single transition state to bubble product. That is related to the normal refolding rate constant

by $(\Delta G^{\pm}_{b,j} - RT\ln\frac{\kappa T}{h}) = -\Delta G^{o}_{hyd,j} = -\Delta C^{o}_{P,j}\left[\left(T - T^{*}_{H}\right) - T\ln\left[\frac{T}{T^{*}_{S}}\right]\right]$ in which the right-hand side is the empirical expression of Murphy et al. Note that since the fitting parameters, T^{o}_{H}, T^{o}_{S}, are both near 385K., Privalov using series expansion of the ln term reduced the latter to $\Delta G^{o}_{hyd,j} \approx T\Delta C^{o}_{P,j}\left(\frac{T-385}{T}\right)^{2}$ with some as yet not established dependence of the heat-capacity change on temperature. As already observed any equations such as these equations that explicitly equate free-energy change and heat-capacity change cannot be correct.

Water is a major participant moving from bulk phase to bubble in amount depending on the temperature but at constant chemical potential of water the associated free-energy change is small. The contributions to enthalpy, entropy and heat-capacity changes are large. Much the same can be said of the charge system of the protein and solution buffer. The above expression also explains why in the collection of mesophilic proteins there is such excellent linear compensation behavior between the pairs of standard hydrophobic enthalpy change and standard hydrophobic entropy change (cf. Fig. @). The cancellation can be shown to be nearly complete in a simple way using compensation theory by plotting the pairs $\Delta H^{o}_{i}, \Delta S^{o}_{i}$ as a compensation plot, Fig. @. The basis for this is the linear-free-energy relationships $\begin{matrix}\Delta G^{o}_{comp,i} = \alpha_{comp} + f_{comp}(i)g_{comp} \\ \Delta G^{o}_{hyd,i} = \alpha_{hyd} + f_{hyd}(i)g_{hyd}\end{matrix}$. The first is

already established using Pohl's results. The second is an assumption to be tested with the further assumption that $f_{comp}(i) = f_{hyd}(i) = i$. Like all "linear free-energy relationships" these are extrathermodynamic since the scaling with respect to protein variation cannot be exact but they are adequate to rationalize the linear plotting in Fig. @. As is often the situation in protein systems the compensation behavior is as precise as the experimental data.

The next major deduction from Privalov et al arises from their finding that the standard enthalpy, entropy and heat-capacity changes in melting of the 13 proteins all scale to the total number of amino-acid residues. They found the relationships $((\Delta H^o_{hyd,i}, \Delta C^o_{hyd.i})\ and\ (\Delta S^o_{hyd,i}, \Delta C^o_{hyd.i})$ linear within error. They are extrathermodynamic relationships due not to the true relationships among temperature derivatives of the LFE equations given above but instead to the fact that the quantities independently scale about linearly with series indes *j*. The LFE are reduced by the compensation relationship between the enthalpy and entropy to the analogs of the two free-energy expressions of those authors.

$$\Delta G^o_{comp,i} = \alpha_{comp}(T) + (T_{c,comp} - T)\Delta S^o_{comp,i}$$
$$\Delta G^o_{hyd,i} = \alpha_{hyd}(T) + (T_{c,hyd} - T)\Delta S^o_{hyd,i}$$

We already know that the compensation temperature for the first of these equations is 354K independent of the number of residues and the temperature and now from the plot of the second equation in Fig. @ using data tabulated by Privalov and Makhatadze we find the compensation plot for the second equation to be linear within small errors and the compensation temperature to be 348K very near the experimental temperature of the data, . Similar plots of their data for temperatures from 298K to 373K establish that within small errors the compensation temperatures are equal to the mean experimental temperatures. At those temperatures $\Delta G^o_{hyd,i} = \alpha_{hyd}(T)$ is a constant approximately given by the

intercepts of the plots and always small with respect to $\Delta H^o_{hyd,i} and\ \Delta S^o_{hyd,i}$. This result generalizes for the common mesophiles Benzinger's deduction for ribonuclease A published in 1971 thus establishing again that folded stability does not depend on interaction of oily sidechains with bulk water. Thus the "hydrophobic-hydration" thermodynamic quantities all refer to bubble formation and measure conformational changes in the motile polypeptide, changes in water content and changes in the water-bubble interface well as pH dependencies and disulfide pairing. These are responsible for the large values of $\Delta H^o_{hyd,i} and\ \Delta S^o_{hyd,i}$ but they make only minor contributors to $\Delta G^o_{hyd,i}$ Dehydration and denaturing agents change the situation dramatically the first prevents bubble formation, the second replaced the bubble product with the a "random-coil" product. It is likely that the large positive heat-capacity change in melting is due primarily to the relaxation of water between bulk states and bubble state.

The deduction from the data for trypsin, trypsinogen, chymotrypsin, chymotrypsinogen, elastase and ribonuclease from Pohl, HEW lysozyme from Segawa and Sugihara and from Murphy et al. papain, staphylococcus nuclease, carbonic anhydrase, cytochrome C, pepsinogen, myoglobin and the K4 fragment of plasminogen is that the activation free-energy change for melting measures the major source of folded stability. As had been show, that is cost of knot disruption. That the work required to destroy the knots is proportional to the number of residues follows the fact that the enthalpy and entropy changes in $\Delta G^o_{comp,j} = \Delta H^o_{comp,j} - T\Delta S^o_{comp,j}$ are normalized for all members of their family by division by the number of residues in each. Murphy et al give the constant enthalpy change per mole of residues as 1530 cal (the same as Pohl's activation enthalpy per mole of residues within small error) and the entropy change per mole of residues as 4.32 cal/K. Hence at any T below 354K the activation free

energy change in melting is $\Delta G_j^{\pm}(T) = n_j(1530 - 4.32T)\text{cal} = \Delta G^o_{comp,j}(T)$ in which n_j is the residue total for protein j, the numerical quantities are per-mole of residue.

The third important deduction following from the second is that the knots of all these proteins as measured by their thermodynamic stability are scaled in terms of their total number of residues. Generally this must mean that they contain the same percentage of the total number of peptide groups but we are not yet able to split of the contributions from disulfide groups. The standard free-energy changes are approximately normalized on division by the number of residues. Note, however, that this argument is considerably oversimplified because pH, solvent and residue-exchange effects are thrown into the hydrophobic free-energy of Murphy et al as are any temperature-dependent effects such as bubble size and water content. In so far as $\Delta G^o_{hyd} << \Delta G^o_{comp}$ all mesophile proteins should melt at 353K and several of the examples from Privalov and Makhatadaze melt near this temperature. Forward and backward rate constants at 354K should be near $\left(\frac{\kappa 354}{h}\right)$. Melting temperatures greater than 354K remain a puzzle possibly to be solved by better understanding of the roles of disulfide bonds.

The characteristic and apparently universal knot temperature of 354K is consistent with the ratio of the numbers of knot residues to the number of matrix residues as a fixed number. One important consequence is that the mesophiles will tend to have the same thermodynamic stability. As we have seen, this is found for the melting rate constant but it is also true to a first approximation for the standard free-energy change in melting. These vary because of differences in hydration but show only minor sensitivity to the number of disulfide groups

Studies specifically designed for numerical characterization of the contracted states of matrices are still of very limited number for reasons already given most often the result of the fact that the conformational changes in matrices are hidden in the coordinate errors of diffraction studies. Calorimetric studies of melting of enzymes with covalent inhibitors such as the acyl enzyme derivatives of the serine proteases are usually for contracted matrices and thus not correctly comparable with those on free proteins. The conformer difference appears to lie between 20 and 40 kcal/M for mesophilic proteins.but systematic comparisons based on B factors though essential have not been carried out.

The activation enthalpy is an estimate of the potential energy required to break the knot and an estimate of the number of residues in a knot. For example, in chymotrypsin the fraction of total residues is

$\frac{\Delta H_f^{\pm}}{1529} = 55$ residues for the fraction $\frac{55}{229} = 0.24$ using Pohl's value of 84kcal/mole for the activation enthalpy. This protein has two knots with 28 residues per knot and thus 14 strong peptide-peptide hydrogen bonds per knot. The number of non-hydrogen atoms per knot is 0.12 x 1600 =192 but only 28 are the N and O atoms of the special hydrogen bonds. This gives an estimate of 12 % of total peptide groups per knot for all proteins in the mezophile class. Higher percentages are suggested for real hyperthermal proteins and archaea

The activation entropy change in melting is somewhat complicated by changes in hydration. Thus the amount of Kuntz "non-freezing water", the B shell hydration, changes with temperature and solvent composition. Total drying increases the standard free-energy change and great reduces the standard enthalpy change. The melting temperature is considerably increased. These effects are due to removal of the native hydration shell and the prevention of normal bubble formation. However, the activation entropy wet or dry signals the general lowering of matrix vibrational frequencies on release of tension.

Morozov and Morozov established that protein mechanical stability disappears at 354K. They measured the Youngs modulus as a function of temperature under normal hydration conditions and found that a deformation enthalpy per residue of 1720 cal per mole and an enthalpy of deformation of 4.9 cal/K per mole of residues was required at 298K to destroy the mechanical stability. These measure the stress that must be added to the knots and thus measure the stress lost between 298K and 354K by increased temperature. The knots with zero stability at 354K have increased stability of $n_i g(298K) = n_i(1720 - 298x4.9) = n_i 260 cal/mole$ at 298K in which n_i is the number of residues in the total protein. Thus for a protein with a total of 245 residues the free energy required to break the knots at 298K is 63.5 kcal. The difference between that values and the standard free energy of bubble formation from the native state is at least roughly that from transition state to bubble including relaxation of matrix stress. In dry proteins the stress is only partially released so melting does not much favor knot disruption and mechanical stability is retained up to 600K.

The agreement among these numbers from the several kinds of experiments establishes the remarkable convergence in protein evolution among the mesophiles. Just as the discovery in evolution of the knot-matrix construction led to a single construction principle for enzymes and to the tailoring of matrices that makes possible the powerful specificity features found in matrix contraction, so it now appears that the common indexing of proteins by their residue number also introduces great simplicity where very little has been expected. It also reveals a useful way to organized protein research since *for each number of total amino-acid residues there is a "standard" protein against with which all modifications and variations of proteins of that size can be compared. to reveal effect and cause.* The existence of this concept immediately greatly simplifies proteins research since it provides a quantitative basis for describing and comparing the invariant

features of knot construction already detected. However despite the fact that the number of residues is the central parameter in such computations knot thermodynamics and melting rates give little information about matrices. However they organized matrix research just as they provide the physical skeleton on which to assemble, sort and understand matrix data such as that from SDM experiments.

The "standard properties" can be defined in several ways but for predominant aqueous solutions of proteins adhering to the two-state model for melting the standard enthalpy change in melting is equal to the activation enthalpy corrected for the RTln (kT/h) rate-theory pre-exponential. The critical temperature for melting is 354K The standard entropy change and the activation entropy for melting are the same. Knot strength as measured by loss of mechanical stability can be computed at any temperature from the enthalpy and entropy parameters provided by Morozov and Morozov.

Just how well our deductions for mesophiles explain true thermophiles and the "extremophiles" from archaea remains to be seen and there is not yet reason to expect those broad classes to display the same pragmatic rules. However, the DNA-binding protein of archaea protein which stabilizes coiled DNA against heat damages as do histones in mammals contains one three-strand and one two-stand β structures fused to fit into major and minor grooves. It has no disulfide bonds and considerable thermal stability. The

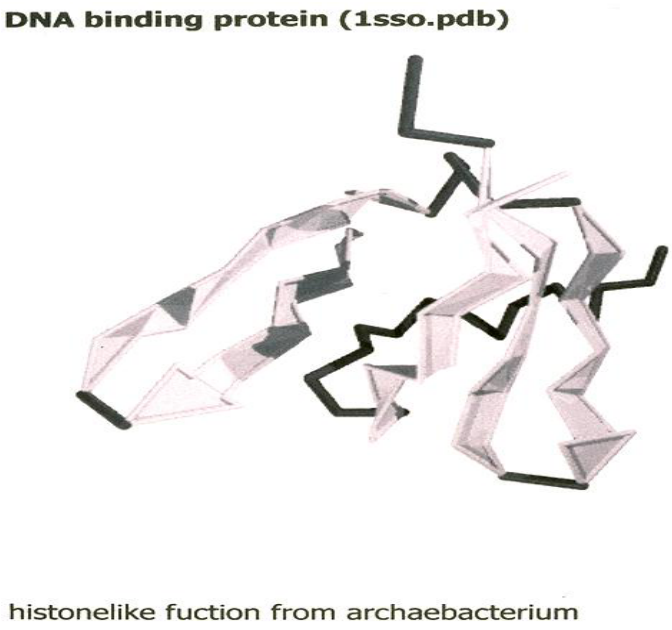

Figure 1 Bacterial DNA binding protein (is it an archaeon?)

similarity between this protein and the G protein of streptococcus, a non-archae protein, suggested by knot contraction and knot rigidity implies that the structural differences are quantitative and not qualitative. The ribozymes

+ thus far characterized appear to be very similar to enzymes in having two hard arms set in soft matrices closely resembling the knot palindromes. It will not be surprising to find that protein catalysts learned their tricks from RNA.