ELSEVIER

Biophysical Chemistry 101–102 (2002) 93–98

Biophysical Chemistry

www.elsevier.com/locate/bpc

# Parsimony in protein evolution

Rufus Lumry*

*Chemistry Department, University of Minnesota, 207 Pleasant St. S.E., Minneapolis, MN 55455, USA*

Received 8 January 2002; accepted 4 March 2002

**Abstract**

Pohl found the activation enthalpy and entropy for melting of his mesophilic proteins to be linear in the total number of residues and Privalov and colleagues found this same linearity for the standard heat-capacity, enthalpy and entropy changes in the overall melting equilibria. Despite the small samples these results suggest that mesophiles individually, and as a class, are related through a single standard representative. If so, very extensive convergent evolution has provided both great simplification and very sophisticated goals for genome decoding and quantitative description of protein substructures [R. Lumry, The protein primer, http://www.umn.edu.chem. /groupslumry].


*Keywords:* Protein evolution; Mesophiles; Enthalpy; Entropy

## 1. Introduction

In our first contribution to this issue [2] we showed how the combination of Pohl's compensation relationship for the melting rate of mesophiles with the analysis of melting equilibrium by Murphy et al. [4] unambiguously establishes that folded stability of mesophiles is due to the loss of electrostatic potential energy as the contraction process moves down the reaction coordinate from the single transition state to the native protein. The second major deduction from these references is that in the absence of complications such as denaturing agents and subzero temperatures the activation thermodynamics scale to the total number of amino acid residues in a protein. Pohl [3] first noticed this for the melting rates of his small collection of serine proteases and then Murphy et al. [4] using the data tabulated by Privalov and Makhatadze [8] showed that the standard enthalpy, entropy and heat-capacity changes in melting of 13 more mesophilic proteins are normalized to a common reference value on division by the total number of amino-acid residues in each protein. The compensation relationships for their data are essentially linear because the standard enthalpy and entropy are so much larger than the standard free-energy changes That behavior is extra-thermodynamic and only indicates that the thermodynamic quantities scale to changes in molecular details in about the same way [5]. However, that scaling is the basis of the dependence on numbers of residues and it is consistent with reasonable similarities among the proteins.

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

0301-4622/02/$ - see front matter 
PII: S0301-4622(02)00189-8

Using standard heat-capacity change as a criterion, Murphy et al. [4] divided the total standard free-energy changes into a part with all the heat-capacity dependence attributed by them to interactions between melted product and bulk water, 'hydrophobic interaction' and a remaining part attribute to polar changes in melting. The corresponding LFE relations are:

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

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

In a previous paper it is shown that the first part of the above equation measures the formation of the transition state from the native state and the second part represents the final step from transition state to micelle-like product so we already know that the compensation temperature for the first of these equations is 354 K—independent of the number of residues—and therefore the temperature is no longer a factor at 354 K. The second had small positive values at all temperatures and since actual unfolding under normal conditions is minimal in denaturation of the mesophilic proteins, the 'hydrophobic-hydration' assumed by Murphy et al. must be due instead to conformational rearrangements and uptake of water.

## 2. Deconstructing proteins

The deduction from the data for trypsin, trypsinogen, chymotrypsin, chymotrypsinogen, elastase and ribonuclease from Pohl [3], HEW lysozyme from Segawa and Sugihara [6], and papain, staphylococcus nuclease, carbonic anhydrase, cytochrome *c*, pepsinogen, myoglobin and the K4 fragment of plasminogen from Murphy et al., is that the activation free-energy change for melting measures the major source of folded stability, that is the work required to destroy the cooperative integrity of the small substructures called knots [7]. Murphy et al. calculated that after normalizing to constant residue number the enthalpy and entropy changes in $\Delta G^{\circ}_{\mathrm{comp},j} = \Delta H^{\circ}_{\mathrm{comp},j} - T\Delta S^{\circ}_{\mathrm{comp},j}$ were reduced to 1530 cal per mole of average residue (the same as Pohl's activation enthalpy per mole of residues) and the entropy change per mole of residues to 4.32 cal/K. Hence, at any $T$ below 354 K the activation free energy change in melting is

$$\Delta G^{\pm}_{j}(T) = n_j(1530 - 4.32\ T)cal = \Delta G^{o}_{comp,j}(T)$$

where $n_j$ is the residue total for protein $j$, the numerical quantities are per-mole of residue. Forward and backward rate constants at 354 K should be near $\left(\frac{\kappa 354}{h}\right)$.

The activation enthalpy for melting is an estimate of the work required to break the knots and an estimate of the number of residues in the knots. For example, in chymotrypsin the fraction of total residues is

$$\frac{\Delta H^{\pm}_{f}}{1529} = 55 \text{ residues for the fraction } \frac{55}{229} = 0.24$$

using Pohl's value of 84 kcal/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 $0.12 \times 1600 = 192$ 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 mesophile class. Higher percentages are suggested for real hyperthermal proteins and archaeons

The standard free-energy changes are approximately normalized on division by the number of residues but variations are produced because pH, solvent and residue-difference effects are thrown into the product-formation as are any temperature-dependent effects such as bubble size and water content. In so far as $\Delta G^{\circ}_{\mathrm{hyd}} \ll \Delta G^{\circ}_{\mathrm{comp}}$ all mesophile proteins should have a melting temperature of 354 K and several of the examples from Privalov and Makhatadze [8] melt near this temperature. However, a small fraction of mesophiles have melting temperatures as high as 373 K. Bovine pancreatic trypsin inhibitor (BPTI) and ribonuclease A are examples. These proteins have large disulfide populations which remain frozen in the step from native state to transition state for melting in which coordinate changes are fractions of angstroms suggesting that the extra stabilizing is due to the fixed disulfides. In the bubble state disulfide rearrangement is fast, as shown by Huggins et al. [9],

Table 1
The pH dependence of melting of the *G* protein from streptococcus

| pH | $T_m$ (K) | $\Delta G^o_{comp}$ (kcal/M) | $\Delta G^o_{hydro}$ (kcal/M) | $\Delta H^o_{total}$ (kcal/M) | $\Delta S^o_{total}$ (kcal/M) | $\Delta H^o_{comp}$ (kcal/M) | $\Delta S^o_{comp}$ (cal/M) | $\Delta H^o_{hydro}$ (kcal/M) | $\Delta S^o_{hydro}$ (kcal/M) |
|---|---|---|---|---|---|---|---|---|---|
| 2.69 | 332 | 15.3 | −15.0 | 169 | 509 | 238 | 672 | −69 | −163 |
| 2.88 | 337 | 11.4 | −11.6 | 195 | 509 | 238 | 672 | −43 | −93 |
| 3.10 | 339 | 10.2 | −10.0 | 189 | 557 | 238 | 672 | −49 | −115 |
| 4.00 | 349 | 3.2 | −3.0 | 205 | 587 | 238 | 672 | −33 | −85 |
| 5.40 | 353 | 0 | 0 | 238 | 672 | 238 | 672 | 0 | 0 |

Data from Alexander et al. [11]. Notation is that of Murphy et al. [4].

indicating high motility unlikely to have much effect on the free energy

Since Pohl's compensation behavior shows the same dependence on residue number, the knots on which it depends must contain the same percentage of the total number of peptide groups and the universal knot temperature of 354 K requires that the ratio of the numbers of knot residues to the number of matrix residues be a fixed number. The melting rates for chymotrypsin derivatives measured by Hopkins and Spikes [10], as well as those for removal of single disulfide groups, measure changes in fully folded species but the refolding rate constant appears to reflect variations in charge and hydration of the bubble state. This deduction has some generality. We have used it in the previous paper to separate changes in thermodynamics of the two rate constants with temperature for a series of related proteins. In Table 1 it is applied to pH dependencies of the two rates for a single protein.

The melting temperature given in Table 1 as $T_m$ varies with pH, and since it is equal to the value of the ratio $(\Delta H^o_j/\Delta S^o_j)$ at each temperature *j*, that ratio is a convenient estimator of the relative contributions to the standard free energy in melting from the two equations used by Murphy et al. When the ratio equals 354 K the entire free-energy change is due to knot destruction [12]. In Table 1 the 'test ratio' is very close to 354 K except at most acid pH values, therefore, with this protein the data suggest that only extreme ionization changes destabilize the native species. However, the table numbers were computed using the assumption that Pohl's compensation plot does not change with pH changes. This is a critical part of the argument and in the absence of direct rate measurements the assumption is required for pH, solvent and disulfide effects. Fortunately, Pohl in his singularly important Habilitation [3] and additional publications, reported melting and refolding rate constants as a function of pH for trypsin and several different proteins finding the assumption and conclusion valid for those proteins. For trypsin he confirmed its independence on solvents including deuterium oxide and ethanol. His comparisons with other proteins suggest that those conclusions are general in dilute buffers but, of course, many new direct rate measurements are required to put melting on a reliable foundation.

## 3. New levels for protein research

That the collection of mesophilic proteins thus far examined is indexed by residue numbers to a common reference is startling evidence of the power of evolutionary trial and error tested by survival value. How these are manifested by archaea proteins and the immunoglobulins has not as yet been studied much although there may now be enough data to make such studies profitable. The extraordinary selectivity in residue selection has major implications for research as both a guide to goal selection and a limitation. As a guide it points to systematic expansion of the popular methods of residue-substitutions to include all residues in a protein in all combinations until the time when quantitative levels of class importance make reliable sorting possible. As a limit it suggests that even at the simplest level the scientific translation of a genome into its proteins may exceed research resources. The latter gains cre-

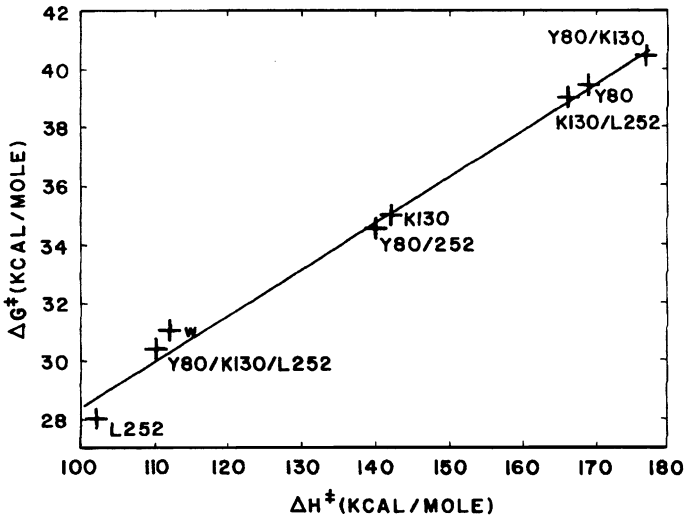


Fig. 1. Compensation plot computed from melting rate date for kanamycin nucleotidyl transferase reported by Dr M. Matsumura, W is wild type,Y80 tyr for asp, K130 tyr for thr, L252 leu for pro. Activation free energy at 298 K. Compensation temperature is 354 K.

dence from additional evidence of the levels of residue selectivity.

In Fig. 1 melting rate data measured by Matsumura with nucleotidyl transferase [23] are shown on a Pohl compensation plot. The residue sites exchanged are all on knots and single exchanges sometimes produce very large changes in both activation enthalpy and activation free energy several of which reveal major increases in knot stability. His study is complemented by that of Kim et al. [13] on the proton-exchange rates and associated stability changes on residue-by-residue substitution of alanyl residues for BPTI knot residues. Both show that only the wild type falls on the 'natural' indexing by residue number. Such residue selectivity taken with a precision in knot construction of 0.05 Å illustrated by the G protein from streptococcus establishes goals not likely to be obtainable without the temperature factors from diffraction studies but there we run into a historic block of much importance. Errors in the length and angle coordinates are still larger than the geometric variations in subtle-change processes but the temperature factors, which do have the necessary precision even in low-resolution studies, are still universally ignored. Protein research is not much different from any other kind of research in so far as once reasonable but casual conclusions have come to retard progress. Proteins are very sophisticated machines improbable, non-chemical devices operating by ad hoc rules discovered by trial and error to produce the catalytic rates and specificity sufficient for biology to exist at all. It is doubtful if any small-molecule chemistry has ever been sufficiently fast and specific to support life. Fortunately the B-value data already exist to terminate the increasingly desperate efforts of many to build biology using small-molecule chemistry.

A number of only slightly less serious misconceptions has crept into the protein literature. Most are described in the Protein Primer [1] but the four causing the most confusion should be mentioned here:

1. Full exposure of polypeptide to bulk water is very limited above the supercooled temperature regime of pure water where only structure breakers such as hydrazine and urea produce it. A classical direct measurement of volume in the normal melted form, more like a motile micelle than even any random-coil approximation, is that of Corbett and Roche [14] who found less

than a twofold increase, a result reported before and since by many others. Evidence for random-coil unfolding in dilute buffers is rare, ambiguous at best

2. The standard enthalpy and entropy changes in normal melting, which in dilute buffers is above 280 K, are positive rather than negative as expected for exposure to bulk water and the positive heat capacity is largely the result of increased conformational freedom in bubble and deviations from the two-state mechanism.
3. A major source of confusion seems to be the 'molten-globule' state which is unlikely to be the new state (postulated to be a critical intermediate in the folding of some proteins) but likely to be either the normal bubble product in thermal denaturation—since that state is rarely correctly factored into denaturation mechanisms—or a half-melted state resulting from out of phase melting of the paired functional domains of proteins, particularly enzymes. As can be found throughout the Protein Database, enzymes always contain two semi-independent proteins well matched in free volumes as indicated by the B-factor palindromes of their knots [15]. Finding abnormal conditions in which one domain melts at somewhat less drastic conditions than the other can generally destroy the good matching necessary for function. Privalov has described stepwise melting of proteins with several functional domains [16] and Brandts et al. [17] developed the quantitative description of linkage between catalytic domains using methanol to effect stepwise denaturation of ribonuclease A. Battistel and Bianchi [18] using hydration changes produced detailed descriptions of the stepwise denaturation of that protein. Volynskaya et al. [19] reported that guanidinium HCl at low concentration produces a small, denatured product in a phase-like process with an 18% increase in volume. They labeled this a molten globule but its characteristics [20] are consistent with those of the normal bubble product formed in pure water.
4. The activation entropy change in melting is still complicated by changes in hydration. In particular the amount of Kuntz 'non-freezing water', the outer hydration shell sometimes called the 'B shell', changes with temperature and solvent composition [21]. At sub-zero temperatures it is a major factor in stabilizing native folding successfully competing against cold denaturation. Total drying at ambient temperatures increases the standard free-energy change and greatly reduces the standard enthalpy change so the melting temperature is much increased. These effects are due to removal of the native-state hydration shell and the prevention of normal bubble formation. The activation entropy for melting wet or dry proteins, signals the general lowering of matrix vibrational frequencies on release of tension. Morozov and Morozov [22] found that Young's modulus of some mesophiles decreases with temperature and is almost negligible at 354 K. The temperature dependence showed a deformation enthalpy per residue of 1720 cal per mole and entropy of deformation of 4.9 cal/K per mole of residues to destroy the mechanical stability at 298 K. Since the values are very nearly the same as those given by Murphy et al. for melting of the knots, they show that the mechanical properties of the whole protein depend on knot strength. Young's modulus measures the stress that must be applied externally to melt the knots. The increase in knot stability on cooling from 354 to 298 K is $n_j(1720-298x4.9)=n_j260$ cal/mole where $n_j$ is the total number of residues in protein *j*. Therefore, for a protein with a total of 245 residues the free energy required to break the knots at 298 K is 63.5 kcal. The difference between that value and the standard free energy of melting is a good total approximation from transition state to bubble including relaxation of matrix stress. In dry proteins the stress is only partially released, therefore, melting does not favor knot disruption and mechanical stability is retained up to 600 K.

## 4. Major implications for protein research

The agreement among these numbers from the numerous experiments reveals a remarkable convergence in the evolution of mesophilic proteins. It is manifested by the indexing of proteins using

their residue numbers, a great simplification where very little has been expected. As shown elsewhere [1,5], the discovery in evolution of the knot-matrix construction led to a single construction principle for enzymes most likely supporting a single 'nutcracker' mechanism using mechanical force for rate activation rather than heat. Similarly tailoring of matrices adjusts both scalar and vector properties of the force vectors generated by matrix contraction so as to produce the powerful specificity features found in catalysis, ligation in general and protein–protein association. Research on these topics will be much more efficient when organized to exploit the simplicity. Thus, *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 numbers given in this article can be used to provide quantitative descriptions of these 'standard proteins' such as knot descriptions, the most promising place to start. Although there are empirical rules found by evolutionary experimentation for placing the knot H bonds, there is nothing simple in the task of finding those rules. Protein families are determined by the placement of the strong knot H bonds and not by any simple residue conservation. Perhaps even more complicated algorithms apply to the selections of matrix residues. The big advantage for those attempting to establish the scientific basis of DNA information and its transformation into protein machines is the fact that only very small sets of residue combinations have been found acceptable in evolution. One can describe the path from genome that must be followed and it is very long: from positioning of knot H-bond units to fixed proportions of knot to residues to pairwise matching of functional domains to a single major mechanism of enzyme function to multi-enzyme with ever increasing cooperativity and so on into deeper and deeper complexity.

How well our deductions stand up for a wider arrange of proteins and newly studied enzymes remains to be seen and there is not yet reason to expect hyperthermal proteins and the archaea proteins to display the same pragmatic rules.

## Acknowledgments

This work was supported by the Lumry Foundation of Hunts Point, Bellevue, Washington.

## References

[1] R. Lumry, The protein primer, http://www.umn.edu.chem. /groupslumry.
[2] R. Lumry, this journal, 2002.
[3] F. Pohl, Habilitationsschrift 'Kinetics of the reversible denaturation of globular proteins'. Universities of Gottingen and Constanz, 1969.
[4] K. Murpy, P. Privalov, S. Gill, Science 247 (1990) 559.
[5] R. Lumry, Methods in Enzymology, chapter 29 259 (1995).
[6] S. Segawa, M. Sugihara, Biopolymers 23 (1984) 2473.
[7] R. Gregory, R. Lumry, Biopolymers 244 (1985) 301.
[8] P. Privalov, G. Makhatadze, Abv. Protein Chem. 47 (1995) 305.
[9] C. Huggins, D. Tapley, E. Jansen, Nature 167 (1951) 592.
[10] T. Hopkina, J Spikes, Biochem. Biophys. Res. Commun. 28 (1967) 480.
[11] P. Alexander, S. Fahnestock, T. Lee, J. Orban, P. Bryan, Biochemistry 31 (1992) 3597.
[12] Ref. 1, Chap. 12.
[13] K-S. Kim, J. Fuchs, C. Woodward, Biochemistry, (1993), 9600.
[14] R. Corbett, R. Roche, Biochemistry 23 (1984) 1888.
[15] R. Lumry, in: R. Gregory (Ed.), Protein-Solvent Interactions, chapter 1, Marcel Dekker, New York, 1995.
[16] P. Privalov, Adv. Prot. Chem. 35 (1982) 1.
[17] J. Brandts, C. Hu, L-N. Lin, M. Mas, Biochemistry 28 (1989) 8588.
[18] E. Battistel, D. Bianchi, J. Phys. Chem. 98 (1994) 5368.
[19] A. Volynskaya, E. Kasumov, A. Shishkov, Molec. Biol. (in Russian) 32 (1998) 377.
[20] M. Schrier, A. Ying, M. Ross, E. Schrier, J. Phys. Chem. 81 (1977) 674.
[21] Rf. 1, Chp. 13.
[22] V. Morozov, T. Morozov, J. Biochem. Struct. Dynam. 11 (1993) 459.
[23] M. Matsumura, S. Sasumura, S. Alba, Nature 323 (1986) 356.