ELSEVIER

Biophysical Chemistry 101–102 (2002) 81–92

Biophysical Chemistry

www.elsevier.com/locate/bpc

Protein substructures and folded stability

Rufus Lumry*

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

Received 4 March 2002; accepted 4 March 2002

Abstract

Protein substructures detected in proton-exchange experiments can be described in quantitative detail with the Debye–Waller temperature factors from diffraction studies. The smallest substructures, in mesophilic proteins approximately 12% of the total residues, determine thermodynamic as well as kinetic stability by electrostatic synergism of a few tightly packed clusters about central peptide–peptide hydrogen bonds. Fixed positions of the clusters establish genetic stability of a protein family. The normal product of thermal denaturation above 280 K in dilute buffers, a compact but motile bubble, is formed with positive free-energy change in step from native state the single transition state and smaller negative change in the step from transition state to product. The largest substructures, approximately 80% of the residues, undergo changes in atom free volumes in function that are small relative to coordinate errors in protein diffraction studies but nevertheless describe the most important conformation changes. The criterion of precision in protein construction is approximately 0.05 Å and may be found to be smaller when precision in X-ray diffraction improves. The ratio of residues in the two substructures is fixed in mesophiles.

Keywords: Protein substructures; Thermodynamic stability; Hydrophobic hydration; X-ray temperature factors

1. Introduction [1]

In dilute buffers the activation heat capacity for the melting rate for mesophilic proteins, the large class stable at ordinary temperatures and below 373 K, has been found to be near zero. Almog et al. [2] showed that increasing concentrations of structure breakers like urea first produce the normal product in dilute buffers then force unfolding to conformations resembling random coils. In between the two-state model no longer applies and the heat-capacity behavior is correspondingly anomalous [3]. Eisenberg and Schwert [4] first demonstrated zero activation heat-capacity in their pioneering work on chymotrypsinogen to be followed 20 years later by Pohl [5] in his measurements of melting rates for ribonuclease A and several members of the trypsin and trypsinogen families in free and acyl forms. Sugihara and Segawa [6] added HEW lysozyme to the list. For all these proteins their melting rate measures the rate of disruption of the small substructures as first indicated by the 353 K slope of Pohl's linear compensation plot of activation enthalpy versus activation entropy changes [7]. Morozov and Morozov [8] found that the Young's modulus for myoglobin, HEW lysozyme, β-lactoglobulin and F-actin threads decreases to a near-zero value at

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

0301-4622/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0301-4622(02)00156-4

Table 1
Variation of the melting parameters for ribonuclease A

T_m K	$\Delta G_f^{\mp}$ kJ/M	$\Delta G_{back}^{\mp}$ kJ/M	ΔG_{total}^{o} kJ/M	$\Delta H_f^{\pm}$ kJ/M	$\Delta S_f^{\pm}$ J/KM	$\Delta H_b^{\pm}$ kJ/M	$\Delta S_b^{\pm}$ J/KM	ΔH_{total}^{o} kJ/M	ΔS_{total}^{o} J/KM
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

Equilibrium data from Makhatadze and Privalov, Table III [13]; Rate data from Pohl [7].

354 K thus mimicking the melting rate constants, as do also the proton exchange rate constants for the small substructures. The large activation free energies for melting make rate measurements without denaturants difficult without special apparatus so their determination has not been routine. Hopkins et al. [9] showed that rates in 8-M urea at pH 7.2 for many chymotrypsin derivatives vary being about thirtyfold slower for the most stable acyl-derivatives such as the diphenyl carbamyl and trimethyl acetyl derivates [10]. The method is limited because urea destroys the B hydration shell before melting can occur but their results show that the derivatives modified the native state rather than transition state or the melted state. Melting rates are essential in measuring thermodynamic stability and as shown below often instead of direct measurement they can be computed from standard thermodynamic changes in melting.

2. Melting rates and equilibrium measurements of protein stability

Pohl found that the values for activation enthalpy and activation entropy vary with residue number and because of their mutual compensation the activation free energy is linearly related to the activation enthalpy. The linearity factor decreases with temperature. It is constant at any temperature for all of his mesophilic proteins and goes to zero for the group when the experimental temperature equals the compensation temperature. Thus,

$$-RT\ln\left(\frac{hk_{f,j}}{\kappa T}\right)=\Delta G_{f,j}^{\mp}=(\Delta H_f^{\pm}-T\Delta S_f^{\pm})$$

$$\frac{\Delta G_{f,j}^{\pm}}{\Delta H_{f,j}^{\pm}}=1-\frac{T}{T_c};\quad\left(T_c=\frac{\Delta H_{f,j}^{\pm}}{\Delta S_{f,j}^{\pm}}\right).$$

The right side is independent of protein, *j*, since the compensation temperature, T_c, is independent of *j*. Thus, judging from the proteins he studied, the compensation temperature is the thermal stability limit for mesophiles. A few exceptions; e.g. bovine pancreatic trypsin inhibitor (BPTI) and ribonuclease A melt up to 373 K and require additional consideration (vide infra).

His results are consistent with the two-state behavior usually indicated by the ratio near unity of van't Hoff to direct calorimeter values for the standard enthalpy of melting [11]. More extensive evidence for the consistency is found by combining Pohl's compensation behavior with an analysis of the thermodynamic data given in 1990 by Murphy et al. [12]. They used standard thermodynamic quantities for equilibrium melting tabulated by Makhatadze and Privalov [13] most tested for adherence to the two-state formal mechanism. The major assumption in their analysis is that the standard heat-capacity change in melting is due to exposure of protein in melted state to bulk water. Murphy et al. divided the total standard free-energy change into a part independent of the heat capacity $\Delta G_{comp,j}^{o}=\Delta H_{comp,j}^{o}-T\Delta S_{comp,j}^{o}$ and a part including all the heat-capacity change, $\Delta G_{hyd,j}^{o}=\Delta C_{P,j}^{o}[(T-T_H^*)-T\ln[T/T_S^*]]$. The parameters in their second equation were determined empirically as equal to 385 K within small errors; *j* is the series index. The data treated in this way support several important deductions of which the first is that the step from transition state to product as fitted to their 'hydrophobic hydration' equation above 280 K is

always negative for T less than 385 K. In contrast to current opinion, one which these authors may still share (but see Privalov Ref. [14]), their analysis shows that incorporation of water in that step favors product formation so that native-state stability is diminished rather than stabilized by interactions of product with environment including hydrophobic hydration. Folded stability then depends entirely on the heat-capacity-independent part of the total standard free-energy change.

So long as the two-state model is an accurate fit to the melting equilibria the procedure of Murphy et al. can be correctly formulated by expanding the forward and backward rate constants into Absolute rate theory expressions:

$$K=\frac{k_{\mathrm{f}}}{k_{\mathrm{b}}}=\frac{\dfrac{\kappa T}{h}\mathrm{e}^{-\Delta G_{\mathrm{f},j}^{\mp}/RT}}{\dfrac{\kappa T}{h}\mathrm{e}^{-\Delta G_{\mathrm{b},j}^{\mp}/RT}};$$

$$\begin{aligned}\Delta G^{\mathrm{o}}&=\left(\Delta G_{\mathrm{f}}^{\pm}-RT\ln\frac{\kappa T}{h}\right)-\left(\Delta G_{\mathrm{b}}^{\pm}-RT\ln\frac{\kappa T}{h}\right)\\&=\left(\Delta H_{\mathrm{f}}^{\pm}-T\Delta S_{\mathrm{f}}^{\pm}-RT\ln\frac{\kappa T}{h}\right)\\&\quad-\left(\Delta H_{\mathrm{b}}^{\pm}-T\Delta S_{\mathrm{b}}^{\pm}-RT\ln\frac{\kappa T}{h}\right)\end{aligned}$$

If these replace the empirical separation of Murphy, the first expression on the right is the conventional rate-theory expression for the forward rate constant and can be identified with their heat-capacity-independent equation by comparison of the two: $\Delta H_{\mathrm{f},j}^{\mp}=\Delta H_{\mathrm{comp},j}^{o}$ and $\Delta S_{\mathrm{f},j}^{\mp}+R\ln\kappa T/h=\Delta S_{\mathrm{comp},j}^{o}$. However, the physical validity of the identifications is established only because it is identical with Pohl's compensation equation for direct measurements of the melting rate adjusted for the factor from the rate theory. In both formulations the melting-rate data give linear compensation behavior with the characteristic compensation temperature of 354 K. 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 (16 kcal/M at 298 K). That compensation temperature has already been identified as characteristic of the small substructures Gregory and Lumry called knots [25], thus validating the computation of the rate theory parameters from the thermodynamic changes in melting equilibria, an unusual but very useful procedure subject to only small errors from ad hoc variations such as charge and disulfide variations in the melted state.

In this unusual way a clean separation is made between contributions measured by the melting rate constant from native state to transition state and those measured by the following step from product to transition state, in the reverse of the refolding rate back from product

$$\begin{aligned}\left(\Delta G_{\mathrm{b},j}^{\pm}-RT\ln\frac{\kappa T}{h}\right)&=-\Delta G_{\mathrm{hyd},j}^{\mathrm{o}}\\&=-\Delta C_{P,j}^{\mathrm{o}}\left[(T-T_H^{*})-T\ln\left[\frac{T}{T_S^{*}}\right]\right]\end{aligned}$$

in which the right-hand side is the empirical expression of Murphy et al.

Water is a major participant moving from bulk phase into the normal product, a slightly expanded, mobile bubble resembling an unstructured micelle, in amounts depending on the temperature. At constant chemical potential of water the associated free-energy change is small but the contributions to enthalpy, entropy and heat-capacity changes are large. The charge system of the protein and buffer behave the same way so the net standard enthalpy and entropy changes nearly compensate each other. This consideration rationalizes near linearity of the linear-compensation plots from Table 1, one of which is shown in Fig. 1.

The fitting parameters for ribonuclease A in the compilation by Makhatadze and Privalov are given in Table 1. At any one of the temperatures except 278 and 373 K a plot of the standard enthalpy change versus standard entropy change for the 11 complete examples in the Makhatadze–Privalov table is linear with a high correlation coefficient. At each temperature each protein has a different standard free-energy change and thus a different intercept value in its plot but the variations are insignificant against the large enthalpy and entropy

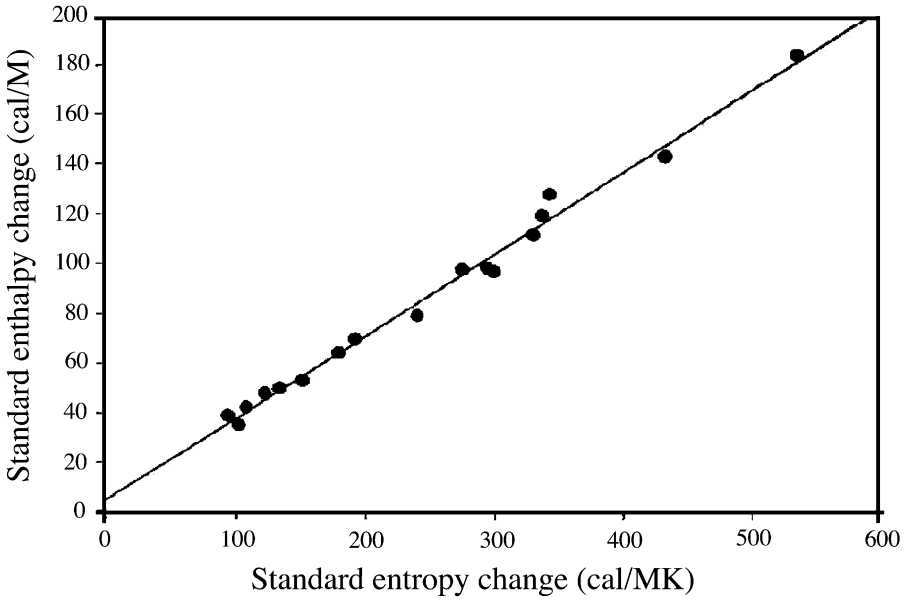

Fig. 1. Linear enthalpy–entropy behavior of a heterogeneous collection of mesophilic proteins illustrated by the data for 323 K. From Table III in Privalov and Makhatadze [13].

changes. At 278 K cold denaturation has become important so there are two products and the linear compensation behavior is lost. The highest temperature 373 K in the table is above the limit of stability for mesophiles and has no physical significance [15].

The standard free-energy contributions from the two steps of Murphy et al. demonstrate independently linear compensation relationships within the precision of the data. Their two equations can be replaced by the LFE expressions

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

so that knowing $T_{\mathrm{c,comp}}$ to be 354 K the values for the $\Delta G^{o}_{\mathrm{hyd},j}$ can be estimated. This procedure is more reliable that use of the heat-capacity expression given by Murphy et al.

The 'hydration' effects responsible for the second equation have contributions from several factors that have not yet been isolated. A major source of confusion is the stress exerted on the knots by the matrices. It is work, that is, free energy, in native species but disappears in the transition state as the rigidity of the native state relaxes. As a consequence the changes in the matrices in response to stress release appear on the product side of the reaction barrier to make major contributions to standard enthalpy and entropy but very little to standard free energy. The osmotic and water-relaxation contributions also have minor free-energy consequences. It is not surprising that the data for Table 1 show extensive enthalpy–entropy compensation behavior and since the outright exposure of peptide to bulk water is minimal in denaturation of the mesophilic proteins, the 'hydrophobic–hydration' assumed by Murphy et al. must be due in major part to conformational factors in forming the small, motile polypeptide product. Those include uptake of water and contract interactions between product 'surface' and bulk water as well as pH dependencies, disulfide pairing and the relaxation of water between bulk and bubble. Sochava [16] has shown that most of the large positive heat-capacity change is due to new modes of motion in the product but the contributions from the relaxation of water

between its bulk states and bubble state cannot be negligible. The new water–peptide associations in the bubble product contribute positive 'hydrophobic' free energy but not the large negative enthalpy and entropy contributions incorrectly suggested by model transfer studies. The latter are due to perturbations of the populations of the two macrostates of bulk water [18]. Dehydration and denaturing agents change the situation dramatically without destroying the cooperativity [17]. The first prevents normal expansion into the normal compact product, the second forces expansion of that product toward true random-coil conformations.

These compensation relationships like all enthalpy–entropy compensation relationships arise from Hammett's famous Linear Free Energy behavior [18,19] and are extra-thermodynamic since the scaling of G, H, S and C_P among members of a family of similar processes (and here with respect to protein variation) cannot be exact. Nevertheless, in protein research the scaling is usually as exact as the data are precise so that where rigorous thermodynamics is overly complicated or inadequate for the non-obligatory linkage systems of living things, extra-thermodynamic relationships provide a useful quantitative theory [20]. Protein denaturation is a good example of their application.

3. Molecular basis of folded stability in mesophilic proteins

The wide separation of exchange-rate groups in the pioneering experiments of Linderstrøm-Lang [21] and their respective coworkers suggested discrete substructures but it was not until 1980 that Gregory et al. [22] extracted the distribution function for the exchange-rate constants to show that most mesophilic proteins have at least three substructures: knots, matrices and surfaces, small groups of sites near protein surface not discussed here. The myoglobin family is an exception with only one. Subsequently it was found that the Debye–Waller factors provide the quantitative descriptions [23]. Those factors routinely tabulated in the Protein Databank under the names 'temperature factor' or 'B factor' have been treated as a by-product in diffraction studies of proteins but in fact since they estimate the free volumes available to atoms with high precision, they contain the most useful and precise molecular information. These factors, available in abundance from modern protein crystallography research, have an experimental error range equal to the range of coordinate changes in the conformation processes that support protein functions, that is, a few tenths of angstroms. As a result they make possible coordinate descriptions of those processes that are quantitatively reliable.

An atom B factor is proportional to the mean square amplitude of the fluctuation of the atom from its idea lattice point computed on the assumption of isotropic scattering. It can be converted to the corresponding estimate of atom free volume or mean displacement fluctuation.

Major uses of B factors are illustrated by Fig. 2, a plot of atom B values against residue number for the G protein from streptococcus [24]. The four atom groups with lowest B values are those responsible for Pohl's compensation plot and thus the major determinant of folded stability. The lowest B values in each are those of the interpeptide hydrogen bonds made very stable by electrostatic synergism resulting from the suppression of permanent polarization, thus producing low local dielectric constant and low potential energy [25]. These few groups of protein structures constitute the knot substructures that must be 'untied' for any unfolding to occur. The knot H bonds can be found from B-factor plots such as Fig. 2 or more simply by sorting the B factors to select the N and O atoms with lowest values. Bahar et al. [26] utilize the fact that the smallest local fluctuations are due to tighter packing so have highest frequencies where packing is most complete. They estimate the frequencies with a simple algorithm based on packing computed from X-ray coordinate data and find good agreement with proton-exchange rates. Knot atoms are easily identified by this promising approach but more detailed local and gestalt information about conformations should be obtainable by the B factors. In particular, B factors are required to measure the molecular changes in the matrix process that drives enzymic catalysis.

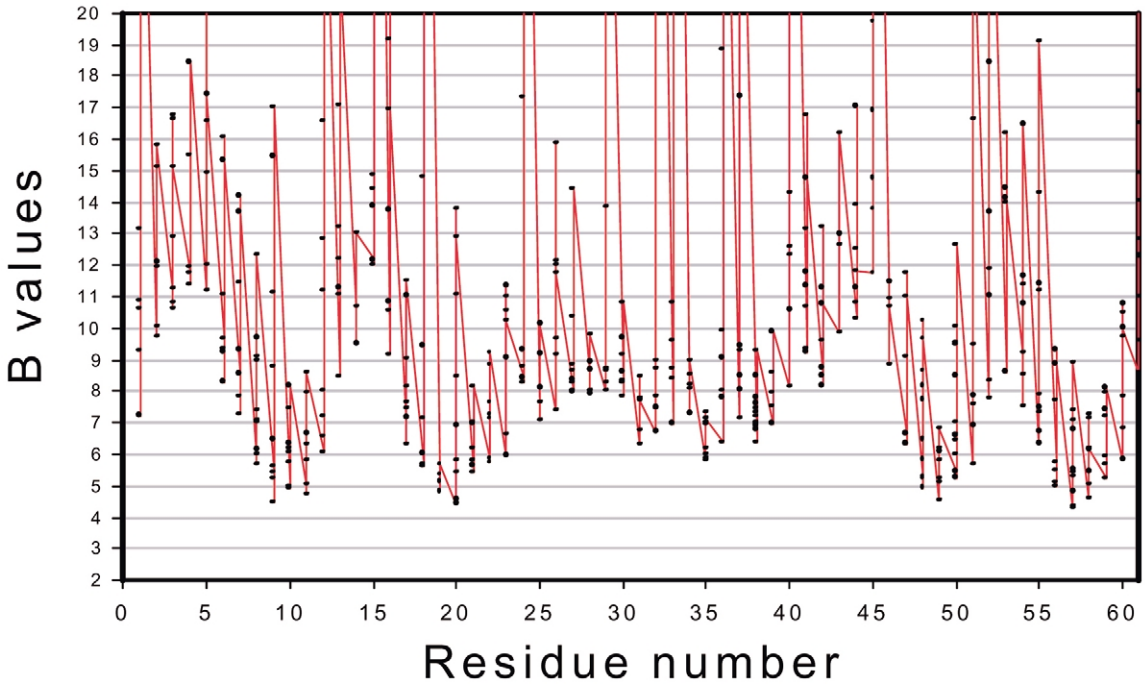

Fig. 2. *B* factors versus residue number for the G protein of streptococcus The atoms with lowest values are the N and O atoms of the peptide–peptide hydrogen bonds formed within the wings and define the palindromic arrangement of the *B* values centered on residue 31, part of a long α-helix. Strong H bonds between the outside edges, those closest to the N and C termini join the two cusps. The non-crystallographic C-2 axis lies in the plane running between the two cusps at residue 31. The knot palindrome pattern generated in this way is one of many different arrangements with C-2 symmetry but somewhat rare among enzymes (phospholipase A-2 is so structured). In this high-resolution X-ray-diffraction study average placement of knot is estimated as 0.05 Å. (PDB file 1lgd; 1 Å resolution).

The knots as patterns of fixed branch points define a protein family thus establishing genetic stability and do so generally without extensive residue conservation. As shown in Table 2 for three members of the trypsin 'family' residue assignments vary somewhat among proteins in the same family often by like-for-like exchanges. The large sequence differences between the halves of the palindrome in each protein do not follow any obvious pattern. Proteins containing only one func-

Table 2
Comparison of knot residues in three members of the trypsin family

N-terminal knot	H-bond partners			C-terminal knot			
Residue number	Chymotrypsin	Trypsin	Elastase	Residue number	Chymotrypsin	Trypsin	Elastase
16	Ile	Ile	Val	228	Tyr	Tyr	Phe
29	Tryp	Tyr	Ser	213	Val	Val	Thr
44	Gly	Gly	Gly	198	Pro	Pro	Pro
56	Ala	Ala	Ala	186	Ser	Glu	Asp
67	Val	Ala	Ala	181	Ile	Phe	Val
85	Ile	Ala	Val	156	Gln	Lys	Gln
102	Asp	Asp	Asp	139	Thr	Ser	Thr

tional domain (single knot with its matrix and surface) do not have the C-2 rotational symmetry closely approximated in enzymes and many other protein classes and probably have different requirements for residue conservation. For example, in Appendix A, the residue conservation in the knots of the BPTI family is shown to be very high. As illustrated by the table, it is a separate functional domain in many larger proteins. The necessity for this unusually high sequence conservation is not yet apparent. Kim et al. [27] found that single-exchanges of knot residues by alanyl residues altered denaturation behavior but only a few exchanges completely eliminated knot formation and cooperativity in melting. Woodward and coworkers were able to melt the matrix by exchanging one non-knot residue 35. Kuroda and Kim [28] extended that study to matrices using multiple alanyl substitutions in a mutant with only one disulfide bond (5–55 between the two ends of the polypeptide). The knot residues 21 and 22 were substituted with little change from wild-type behavior but alanyl residues at matrix positions 26–28 destroyed matrix folding. As a single functional unit its molten-globule state is a melted matrix with intact knot corresponding to neither native state nor a half-melted 'molten-globule' species. Substitutions retaining the knot inhibited trypsin to various degrees.

The point of summarizing these important studies is to suggest that BPTI may be qualitatively different from enzymes and other kinds of proteins with dynamically paired functional domains so that generalizations from BPTI experiments may be misleading. Hemoglobin and myoglobin have no low-*B* atoms characteristic of knots and in proton-exchange experiments appear as single large matrices. They do, however, show rough dyad symmetry on a fragile (non-Arrhenius) free-energy surface [29,32]. Fragile surfaces are rare for mesophiles and those of the myoglobin-fold proteins may explain the wide range of conformers detected by Frauenfelder et al. [30]. In any event hemoglobin and myoglobin fall into a qualitatively different class and there is evidence from antidigoxin suggesting a still different class for the γ-globulins.

Fig. 2 is a plot of atom *B* factors vs. residue number for the G 24 protein of streptococcus. The atoms with lowest values are the N and O atoms of the peptide–peptide hydrogen bonds central to the two knots. Those bonds are always formed within the wings but between wings only in rare instances when used as a hinge of a domain pair. The knot *B* factors demonstrate a palindromic arrangement of the *B* values centered on residue 31. The symmetry of the pattern is very close to C-2 but probably does not need to be exact (T-1 nuclease, vide infra) knot palindrome pattern in this 1-angstrom resolution X-ray-diffraction study is accurate to approximately 0.05 Å [22].

In mesophiles all but a few percent of the non-knot residues form single substructures in which physiological functions are located and modified by residue mutation. In enzymes these substructures already named 'matrices' [31] undergo major changes in free volume during function up to as much as a 50% decrease. Their free-volume range measured by the *B* values is larger than that of the knots, the proton-exchange rates are much faster with much lower activation energies, their compensation temperature is much higher (450 K as opposed to 354 K [32]). Matrices have poor palindrome patterns or none at all [33]. The matrix of the G protein from streptococcus, a non-enzymic protein, illustrates matrices in proteins with matched functional domains. Residue is picture complete except for a long α-helix with high *B* values is shown in Figs. 3 and 4.

The residues containing the lowest *B* values are colored yellow and the lowest in that group are knot residues. Most of the H bonds shown in yellow are reported by Cornilescu et al. [24] using NMR to have relatively large through-bond (j) coupling possibly indicating some covalent character contributing to the high knot strength. Those authors give hydrogen-bond length estimate from diffraction data as shorter than normal, but no errors are specified. Using two independent methods for hydrogen-bond lengths Harris et al. [34] have estimated coordinate errors for hydrogen bonds between side chains from protein X-ray diffraction work as being in the range 0.2–0.8 Å depending on resolution. However, the *B* factors for knot atoms in Fig. 2 show that the actual

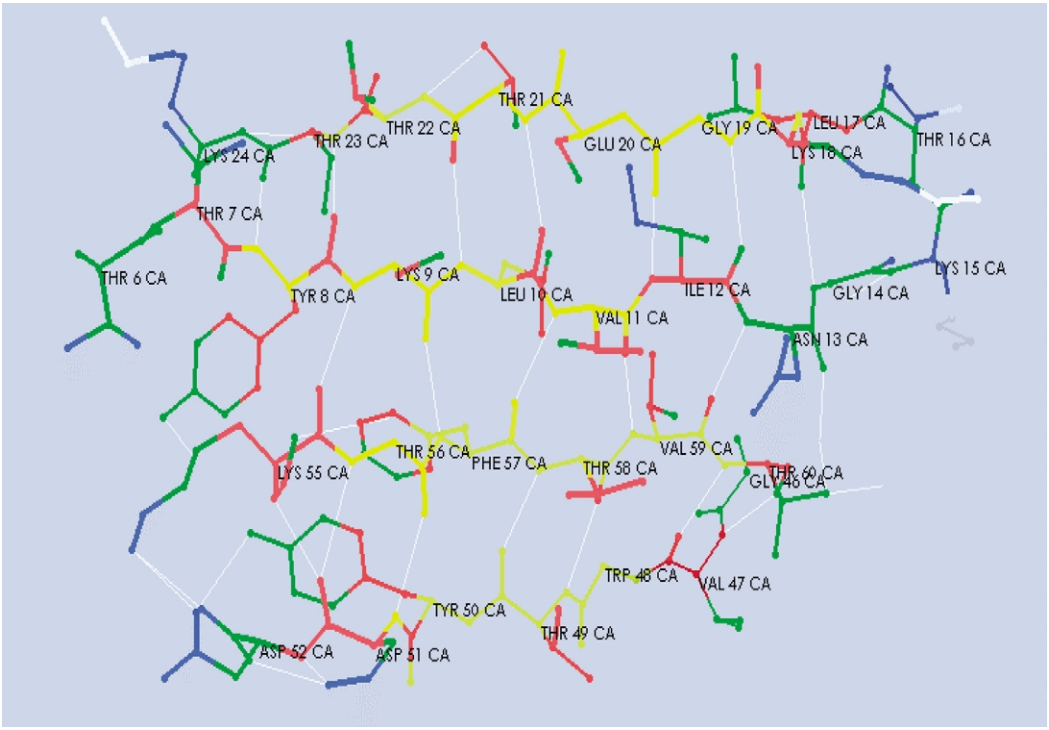

Fig. 3. Backbone picture of the G protein to show the hydrogen-bond array. The yellow regions contain the shorter H bonds found by Cornilescu et al. to have unusually large through bond coupling [24]. The cusps lie above and below the horizontal midline. The C-2 axis of the *B*-factor palindrome is perpendicular to the page plane rising from the center point to pass through a long strap helix running diagonally across this picture (not shown).

coordinate errors are small fractions of an angstrom. As computed from the lowest *B* values in Fig. 2 the error is approximately 0.05 Å. The unusual feature of the construction of the G protein and enzymes in general making the better error estimates possible is the palindrome pattern of *B* values. Scale errors in *B* values not yet clearly understood [35] cancel in comparing the two cusps of the palindrome. That may be the most accurate estimate of the precision of protein construction now available. It is also a good estimate of the standard errors in *B* values. Improving X-ray precision may reduce that value for knot atoms although obviously not much. Recall that chemical events in primary bonds and qualitative changes in hydrogen bonds lie approximately 0.3 Å well within the *B* precision range but well outside the range of errors in X-ray coordinates.

The high precision demonstrated by the *B* factors implies little crystal disorder and small, fast vibrational behavior. Low-temperature data such as those reported for ribonuclease A by Tilton and Petsko show that cooling to 100 K to suppress thermal motion reduces matrix *B* values by approximately 50% with no detectable effect on knot *B* values 32. Protein crystals are strong enough to adjust to those changes without destroying either the crystal or the diffraction pattern. That is very convenient because it means that protein length and angle coordinates do not change much on solution. However, the *B* values undoubtedly do change significantly with the changes in protein activity coefficients. For example, glycerol added to the mother liquor raises the interfacial free energy of HEW lysozyme causing it to contract to lower matrix *B* values [32]. The classic crystalliz-

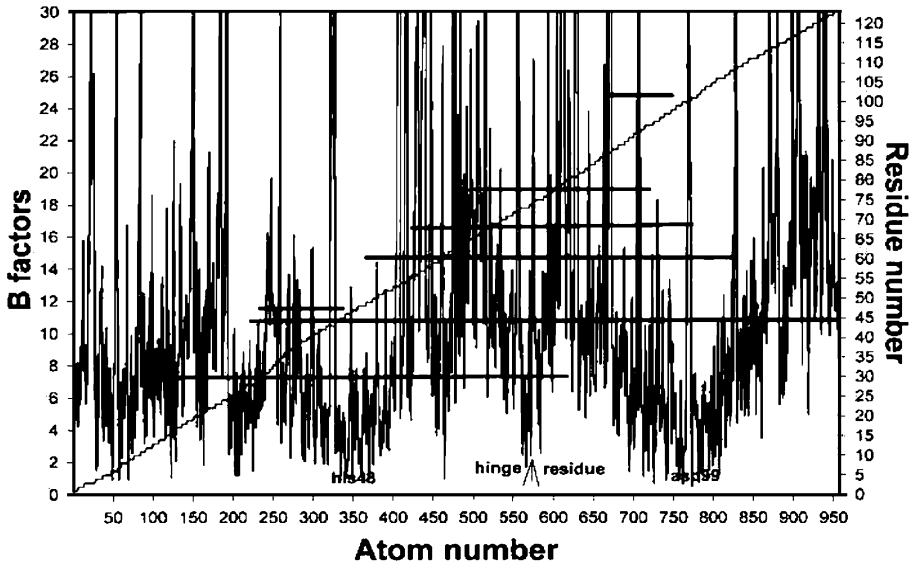

Fig. 4. Phospholipase A2 has two C-2-matched helix cusps each broken at the middle to expose one of the functional groups (His4 or Asp99) the hinge between these functional domains consists of the three disulfide groups (horizontal bars) centered symmetrically on the cusps. Pure helix knots such as found in this enzyme are not common. Lipid substrates lie in a trough between the knots.

ing agents, sulfate ion and polyethylene glycol, are effective because they make free water unavailable at low mole fractions so that they raise protein activities even more [36].

Major deviations from palindromy in knot B factors are very rare but probably very important. T-1 nuclease is an example, the only one we have thus far found Fig. 5. Both knots are large. The α-helix section in the knot in one cusp is balanced by a section of three-stand antiparallel sheet in the other knot. The twofold rotational symmetry characteristic of an exact palindrome is lost but the free volumes of the knots appear to be very well matched. Despite the novelty of this exception it appears to vitiate the exact palindrome requirement for knots suggested by other proteins. The balancing of atom free volumes that remains suggest that matching of the conformation dynamics of the cusps is a better criterion of efficient protein evolution.

The conformation change in the enzyme function now identifiable with the 'subtle' changes first described by Lumry and Biltonen [37] and more recently updated by Gregory [38] occurs in the matrices and consists of cooperative changes in atom free volumes that approximate first-order phase changes despite the fact that they are a result of evolutionary selection rather than any intrinsic property of polypeptides [39]. The systematic increase of B value along residue side chains toward the periphery shown in Fig. 2 suggests how the process has been achieved in some proteins. The amplitude of the expansion–contraction mode of matrices free of specific substrates or inhibitors is roughly 0.5 Å as estimated from the change in matrix B values in intermediate states of catalysis and in binding of specific inhibitors. The period as measured by dielectric dispersion is roughly $(10^{-1}-1)$ ns (3×10^{2} ns at -10 °C [40]). The subtle-change process is a major feature in physiological function but it has not been well characterized. It is accessible through the B factors, circular dichroism, dielectric relaxation, infra-red spectroscopy, fluorescence, phosphorescence and other methods sensing fine adjustments of conformation. It gates the diffusion of catalysts for proton exchange to matrix exchange sites so it is respon-

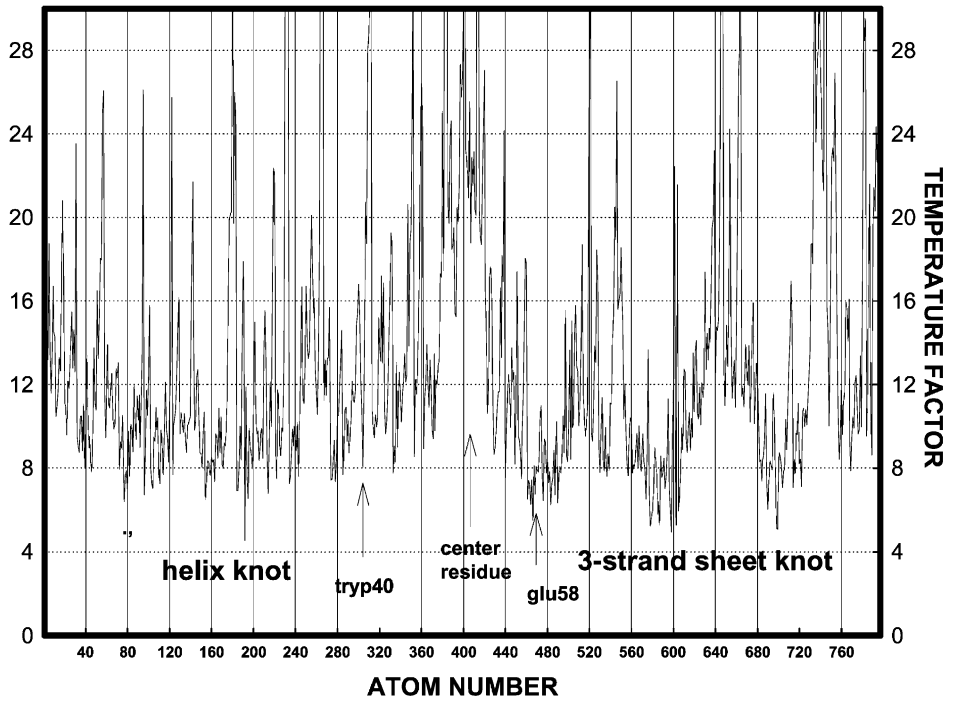

Fig. 5. Atom and residue plots of the atom B factors of T-1 nuclease to show the absence of even approximate palindromy in the knots. Instead the free volume patterns of the knot cusps are closely matched (see text). The two functional groups, Tryp40 and Glu58, are not in the usual matching sequence positions found with good palindrome patterns but form the single interdomain hydrogen bond. (PDB·9rnt).

sible for the preservation in rank-order of proton exchange found by Woodward and Rosenberg [41].

4. Summary

The discovery of protein substructures 20 years ago finally made understanding of conformation and function possible. In this contribution our general explanation of stable folding of mesophilic proteins has been given quantitative description by converting the empirical expressions for free-energy change in thermal melting to the more conventional treatment of the melting rate constant by Pohl. Stability is a consequence of the low electrostatic potential energy of small substructures called knots and is opposed by stress developed in contraction of the large substructures called matrices. In thermal denaturation direct interaction between polypeptide and bulk water in the denaturation product is minimal but the popular assumption that that state has more random-coil conformational character than similarity to micelles has produced much unnecessary confusion. The temperature factors (Debye–Waller factors) from diffraction studies provide the precision necessary for quantitative conversion of structural information into molecular descriptions of structure and function. In the companion paper we further elaborate the data analysis of Murphy et al. to show its implication that all mesophiles can be reduced to a single reference by normalizing the number of amino acids. In a following paper we will show that since all enzymes appear to be have the same essential collection of construction features, there is probably a single catalytic mechanism based on mechanical activation of rate processes rather than thermal activation. A preliminary discussion of these topics is given in *The Protein Primer*.

Acknowledgments

This contribution is a partial up date of Ref. [37] presented at a conference on protein thermodynamics in Kansas City in 1990 held under the auspices NIH, NSF, the Veterans Administration and the Kansas City Science Committee. I am particularly grateful to Professor Roger Gregory with whom the protein substructures were discovered and established. The definitive work on those substructures in Ref. [25]. The Lumry Family Foundation of Bellevue, Washington supports our work. The figures are reproduced from *The Protein Primer* with permission of the copyright owner.

Appendix A:

(1) Note that since the fitting parameters, T_H^o, T_S^o, are both near 385 K. Privalov using series expansion of the ln term reduced the latter to $\Delta G^o_{\mathrm{hyd},j} \approx T\Delta C^o_{P,j}(T-385/T)^2$ with some as yet not established dependence of the heat-capacity change on temperature. As has been shown elsewhere [36] any equations such as these that algebraically equate free-energy change and heat-capacity change cannot be correct. That important deduction by Benzinger [42] is essential in interpreting enthalpy, entropy and volume data from isothermal processes but it is not yet generally known or it is still confused with the functional $\Delta C_P = -T(\partial^2 \Delta G/\partial T^2)$.

(2) Residue conservation in members of the bovine pancreatic trypsin family (BPTI) (Data from the Sequence Database; Top line is residue number in the wild types).

Protein	5	12	14	21	22	23	30	33	35	37	38	43	45	51	55
BPTI	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
Bovine spleen inhibitor	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
Bovine serum inhibitor	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
Turtle egg-white inhibitor	C	G	C	Y	P	Y	C	F	Y	G	C	N	F	C	C
Snail inhibitor K	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
Sea anemone inhibitor	C	G	C	Y	Y	Y	C	F	Y	G	C		F	C	C
Russel's viper inhibitor	C	G	C	I	Y	Y	C	F	Y	G	C	N	F	C	C
Ringhal's cobra inhibitor	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Cape cobra inhibitor	C	G	C	F	H	Y	C	F	Y	G	C	N	F	C	C
Black mamba toxin B	C	G	C	F	H	Y	C	F	Y	G	C	N	F	C	C
Black mamba toxin E	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Black mamba toxin I	C	G	C	F	Y	Y	C	F	W	G	C	N	F	C	C
Green mamba toxin I	C	G	C	F	Y	Y	C	F	W	G	C	N	F	C	C
Black mamba toxin K	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Green mamba toxin K	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Banded krait inhibitor	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Longf-nosed viper CT inhibitor	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Long-nosed viper tryp inhibitor	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Silkworm chymotryp inhibitor	C	G	C	Y	S	Y	C	F	Y	G	C	N	F	C	C
Beta 1-bungarotoxin B chain	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Beta 2-bungarotoxin B chain	C	K	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Inter-alpha tryp inhibitor bovine	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
Trypstatin	C	G	C	L	A	F	C	F	Y	G	C	N	F	C	C
Human A40751 amyloid protein precursor	C	G	C	W	Y	F	C	F	Y	G	C	N	F	C	C
Lipoprotein-associated coagulation inhibitor	C	G	C	F	F	F	C	F	Y	G	C	N	F	C	C
LACI(118–188)	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
LACi(210–280)	C	G	C	F	F	Y	C	F	Y	G	C	N	F	C	C
Alpha3 chain of type VI collagen	C	G	C	W	Y	Y	C	F	Y	G	C	N	F	C	C

References

[1] R. Lumry, The Protein Primer. Available from http://www.umn.edu.chem/groupslumry.
[2] R. Almog, M. Schrier, E. Schrier, J. Phys. Chem. 82 (1978) 1703.
[3] Ref. [1] Chapter 3.
[4] M. Eisenberg, G. Schwert, J. Gen. Physiol. 34 (1951) 583.
[5] (a) F. Pohl, Habilitationsschrift Kinetics of reversible denaturation of globular proteins, Universities of Göttingen and Constanz 1969, FEBS Lett. 3 (609) (1968); (b) Eur. J. Biochem. 4 (373, 7146) (1968); (c) FEBS Lett. 65 (239) (1969).
[6] S. Segawa, M. Sugihara, Biopolymers 23 (1984) 2473.
[7] F. Pohl, Replot of the compensation data in Ref. [36] page 677.
[8] V. Morozov, T. Morozov, J. Biomol. Struct. Dyn. 11 (1993) 459.
[9] T. Hopkins, J. Spikes, Biochem. Biophys. Res. Comm. 28 (1967) 480.
[10] T. Hopkins, et al., Published with permission in: S. Kuby (Ed.), Study of Enzymes, CRC Press, vol. 2, 50.
[11] R. Lumry, R. Biltonen, J. Brandts, Biopolymers 4 (1966) 917.
[12] K. Murphy, P. Privalov, S. Gill, Science 247 (1990) 559.
[13] (a) P. Privalov, G. Makhatadeze, Adv. Prot. Chem. 47 (1995) 305
(b) P. Privalov, S. Gill, Adv. Protein Chem. 39 (1988) 191.
[14] P. Privalov, CRC Crit. Rev. Biochem. Mol. Biol. 25 (1990) 281.
[15] C. Huggins, D. Tapley, E. Jensen, Nature 167 (1951) 592.
[16] I.V. Sochava, Biophys. Chem. 69 (1997) 31.
[17] Ref. [1], Chapter 8.
[18] E. Grunwald, Thermodynamics of Molecular Species, Wiley, New York, 1997.
[19] R. Lumry, S. Rajender, Biopolymers 9n (1970) 1128.
[20] R. Lumry, R. Gregory, in: G.R. Welch (Ed.), The Fluctuating Enzyme, Wiley Interscience, New York, 1986, Chapter 1.
[21] L. Reyerson, W. Hnojweyj, J. Phys. Chem. 64 (1960) 811.
[22] R. Gregory, Biopolymers 22 (1983) 895.
[23] R. Lumry, in: R. Gregory (Ed.), Protein–Solvent Interactions, Dekker, New York, 1995, Chapter 1.
[24] G. Cornilescu, B. Ramirez, M.K. Frank, G. Clore, A. Gronenborn, A. Bax, J. Am. Chem. Soc. 62 (1999) 6275.
[25] R. Gregory, R. Lumry, Biopolymers 24 (1985) 301.
[26] (a) I. Bahar, R. Jernigan, J. Mol. Biol. 266 (1997) 195 (b) I. Bahar, A. Atilgan, M. Demirel, B. Erman, Phys. Rev. Lett. 80 (1998) 2733.
[27] Kim, Fuchs and Woodward. Biochemistry 32 (1993) 9600.
[28] Y. Kuroda, P. Kim, J. Mol. Biol. 296 (2000) 493.
[29] Ref. [1], Chapter 16.
[30] H. Frauenfelder, S. Sligar, P. Wolynes, Science 254 (1991) 1598.
[31] Ref. [1] Chap. 3.
[32] R. Gregory, in: R. Gregory (Ed.), Protein–Solvent Interactions, M. Dekker, New York, 1995, p. 191.
[33] Ref. [1], Chap. 4.
[34] T. Harris, Q. Zhao, A. Mildvan, J. Mol. Struct. 552 (2000) 97. Chapter 4.
[35] T. Singh, W. Bode, R. Huber, Acta Crystallographica, B 36 (1980) 621.
[36] R. Lumry, Methods Enzymology 259 (1995), Chapter 29.
[37] R. Lumry, R. Biltonen, in: S. Timasheff, G. Fasman (Eds.), Structure and Function of Biological Macromolecules, Dekker, New York, 1969, Chapter 2.
[38] 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.
[39] Ref. [1], Chapter 18.
[40] N. Mura, Y. Hayashi, S. Mashimo, Biopolymers 39 (1996) 183.
[41] C. Woodward, A. Rosenberg, J. Biol. Chem. 246 (1971) 4114.
[42] T.H. Benzinger, Nature 229 (1971) 100.