Use of Liquid Hydrocarbon and Amide Transfer Data To Estimate Contributions to Thermodynamic Functions of Protein Folding from the Removal of Nonpolar and Polar Surface from Water†

Ruth S. Spolar,‡ Jeff R. Livingstone,§ and M. Thomas Record, Jr.*,‡,§

Departments of Chemistry and Biochemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706

Received September 9, 1991; Revised Manuscript Received January 24, 1992

ABSTRACT: This extension of the liquid hydrocarbon model seeks to quantify the thermodynamic contributions to protein stability from the removal of nonpolar and polar surface from water. Thermodynamic data for the transfer of hydrocarbons and organic amides from water to the pure liquid phase are analyzed to obtain contributions to the thermodynamics of folding from the reduction in water-accessible surface area. Although the removal of nonpolar surface makes the dominant contribution to the standard heat capacity change of folding (ΔC°_{fold}), here we show that inclusion of the contribution from removal of *polar* surface allows a quantitative prediction of ΔC°_{fold} within the uncertainty of the calorimetrically determined value. Moreover, analysis of the contribution of polar surface area to the enthalpy of transfer of liquid amides provides a means of estimating the contributions from changes in nonpolar and polar surface area as well as other factors to the enthalpy of folding (ΔH°_{fold}). In addition to estimates of ΔH°_{fold}, this extension of the liquid hydrocarbon model provides a thermodynamic explanation for the observation [Privalov, P. L., & Khechinashvili, N. N. (1974) *J. Mol. Biol. 86*, 665–684] that the specific enthalpy of folding (cal g^{-1}) of a number of globular proteins converges to a common value at approximately 383 K. Because amounts of nonpolar and polar surface area buried by these proteins upon folding are found to be linear functions of molar mass, estimates of both ΔC°_{fold} and ΔH°_{fold} may be obtained given only the molar mass of the protein of interest. Use of ΔC°_{fold} and ΔH°_{fold} in conjunction with a melting temperature (T_m) determined under specified solution conditions allows one to estimate the stability (ΔG°_{fold}) under these conditions at any temperature. Calculated values of ΔG°_{fold} in the range 293–320 K appear generally to agree with those determined by calorimetric and noncalorimetric methods. This analysis should provide a basis for predicting the stability of proteins whose structural and thermodynamic properties are similar to those in the data set used for this analysis. Deviations from these predictions may be interpretable in terms of deviations from the average properties of the proteins in the data set.

What interactions are responsible for the stability of folded proteins? Stability is defined thermodynamically as the free energy difference between the native and denatured states: $\Delta G^\circ_{fold} = G^\circ_N - G^\circ_D$. A long-term goal of biophysical studies is the prediction of ΔG°_{fold} and its enthalpic and entropic components for any protein under any conditions (temperature, pH, etc.) of interest from structural data and thermodynamic data on model compounds. Both the enthalpic (ΔH°_{fold}) and entropic ($-T\Delta S^\circ_{fold}$) contributions to ΔG°_{fold} are observed to be highly temperature-dependent, because folding is characterized by a large negative, relatively temperature-independent heat capacity change (ΔC°_{fold}) [cf. reviews by Privalov (1979) and Privalov and Gill (1988)]. The extreme temperature dependence of the enthalpic and entropic contributions to ΔG°_{fold} complicates any attempt at a molecular interpretation of the thermodynamics of protein folding based on protein data alone. To dissect the contributions of individual noncovalent interactions to protein folding, thermodynamic data for the transfer of small organic compounds from water to a nonaqueous phase have been utilized [see, for example, Kauzmann (1959), Tanford (1980), Makhatadze and Privalov (1990), Murphy et al. (1990), and Privalov and Makhatadze (1990)]. These data have been combined with changes in water-accessible surface area in attempts to predict the stability of the native state [see, for example, Chothia (1975), Rashin (1984), Eisenberg and McLachlan (1986), and Ooi and Oobatake (1988, 1991)]. Quantitative predictions of the thermodynamics of folding utilizing the free energy (ΔG°_{tr}) and entropy (ΔS°_{tr}) of transfer require the proper choice of both the standard state (i.e., concentration scale) and the nonaqueous reference phase (i.e., whether transfer is from water to a gas, liquid, or solid). However, quantitative use of the enthalpy (ΔH°_{tr}) and heat capacity (ΔC°_{tr}) of transfer in analyses of the thermodynamics of folding is essentially independent of the standard-state convention, although still dependent upon the choice of nonaqueous phase. The question of the most appropriate standard state convention for the transfer data (Ben-Naim, 1978; Tanford, 1980; Sharp et al., 1991a,b) can be circumvented by working with ΔH°_{tr} and ΔC°_{tr} but not ΔS°_{tr}. If ΔH°_{fold} and ΔC°_{fold} can be estimated from ΔH°_{tr} and ΔC°_{tr} of the appropriate model compounds, then ΔG°_{fold} and ΔS°_{fold} may be evaluated as a function of temperature from one accurate determination of T_m for the solution conditions of interest:

$$\Delta S^\circ_{fold}(T_m) = \Delta H^\circ_{fold}(T_m)/T_m \quad (1)$$

$$\Delta G^\circ_{fold}(T) = [\Delta H^\circ_{fold}(T_m)](1 - T/T_m) + [\Delta C^\circ_{fold}](T - T_m - T \ln T/T_m) \quad (2)$$

[Equation 2 assumes that ΔC°_{fold} is independent of temperature in the range between T_m and the temperature T of interest.

† This work was supported in part by NIH Grant GM23467.

* Correspondence should be addressed to this author.
‡ Department of Chemistry.
§ Department of Biochemistry.

0006-2960/92/0431-3947$03.00/0

The validity of this assumption has been discussed by Livingstone et al. (1991).]

Background and Motivation

In our previous work (Spolar et al., 1989; Livingstone et al., 1991), we found that heat capacity changes for the transfer of hydrocarbons ($\Delta C^\circ_{tr,np}$) and the folding of proteins (ΔC°_{fold}) exhibit similar proportionalities to the area of nonpolar surface removed from exposure to water (A_{np} for hydrocarbon transfer, ΔA_{np} for protein folding):

$$\text{hydrocarbons: } \Delta C^\circ_{tr,np} = -(0.32 \pm 0.08)A_{np} \text{ cal mol}^{-1} \text{ K}^{-1} \quad (3)$$

$$\text{proteins: } \Delta C^\circ_{fold} = -(0.25 \pm 0.03)\Delta A_{np} \text{ cal mol}^{-1} \text{ K}^{-1} \quad (4)$$

[In eq 4, ΔA_{np} (Å^2) is the magnitude of the difference between the water-accessible nonpolar surface areas of the denatured and native state of the protein. The uncertainties cited represent two standard deviations (SD).] Consequently, we proposed that ΔC°_{fold} could be modeled quantitatively using the transfer of hydrocarbons from water to the pure liquid phase. However, a comparison of enthalpies of protein folding (ΔH°_{fold}) and enthalpies of transfer of liquid hydrocarbons (ΔH°_{tr}) has led to criticism of the use of liquid hydrocarbons as a model system (Murphy et al., 1990). [For a recent comment on this question, see Herzfeld (1991).]

For a process with a temperature-independent heat capacity change (ΔC°), the enthalpy change $\Delta H^\circ(T)$ at temperature T may be expressed

$$\Delta H^\circ(T) = (T - T_H^*)\Delta C^\circ + \Delta H^\circ(T_H^*) \quad (5)$$

where T_H^* is a convenient reference temperature. [Note that T_H^* in general differs from the previously defined characteristic temperature T_H where $\Delta H^\circ = 0$ (Baldwin, 1986; Lee, 1991).] For at least some sets of similar compounds (e.g., liquid hydrocarbons, solid cyclic diamides) undergoing the same process, a plot of ΔH° versus ΔC° at a given temperature is observed to be linear, demonstrating the existence of characteristic temperatures T_H^* where ΔH° is the same for each compound in the set (Murphy et al., 1990). For the specific case of transferring hydrocarbons from the infinitely dilute aqueous solution to the pure liquid phase, Baldwin (1986) noted that, at $T_H^* = T_H = 295 \pm 5$ K, the enthalpy of transfer ($\Delta H^\circ_{tr,np}$) is zero:

$$\Delta H^\circ_{tr,np}(T) = (T - 295)\Delta C^\circ_{tr,np} \quad (6)$$

Consequently, Baldwin proposed that the contribution to ΔH°_{fold} arising from the removal of nonpolar surface from water ($\Delta H^\circ_{fold,np}$) could be estimated from ΔC°_{fold}:

$$\Delta H^\circ_{fold,np}(T) = (T - 295)\Delta C^\circ_{tr,np} \cong (T - 295)\Delta C^\circ_{fold} \quad (7)$$

Spolar et al. (1989) and Livingstone et al. (1991) provided evidence in support of this proposal (the liquid hydrocarbon model) based on the similarity of the dependences of $\Delta C^\circ_{tr,np}$ and ΔC°_{fold} on nonpolar surface area (cf. eqs 3 and 4).

Using the liquid hydrocarbon model, the enthalpy of protein folding can be dissected into a contribution from the burial of nonpolar surface and all other contributions (ΔH°_{other}):

$$\Delta H^\circ_{fold}(T) = \Delta H^\circ_{fold,np}(T) + \Delta H^\circ_{other}(T) = (T - 295)\,\Delta C^\circ_{fold} + \Delta H^\circ_{other}(T) \quad (8)$$

Equation 8 predicts that if ΔH°_{other} is independent of temperature, then ΔH°_{fold} and ΔC°_{fold} are linearly related with a slope of $(T - 295)$. However, a plot of this type shows no correlation (linear or otherwise) between the *molar* quantities ΔH°_{fold} and ΔC°_{fold} at 298 K. On the other hand, Murphy et al. (1990) demonstrated that *specific* enthalpy of protein folding (Δh°_{fold}, cal g^{-1}) is a linear function of the *specific* heat capacity of protein folding (Δc°_{fold}) for a given set of proteins. The slope of this plot yields the characteristic temperature $T_H^* = 383$ K [cf. Privalov and Khechinashvili (1974)], which is very different from the liquid hydrocarbon prediction (295 K). On the basis of this observation, Murphy et al. (1990) concluded that liquid hydrocarbon transfer data do not adequately model the thermodynamics of protein folding.

However, another possible explanation for the apparent discrepancy between the liquid hydrocarbon and protein folding data is that ΔH°_{other} (eq 8) is temperature-dependent, i.e., that ΔC°_{fold} is not entirely determined by the burial of nonpolar surface but may also include a detectable contribution from the burial of polar surface. To refine our model of protein folding based on the transfer of model compounds, we have investigated whether the removal of *polar* surface area (ΔA_p) from water contributes to ΔC°_{fold} and ΔH°_{fold} and whether its inclusion can explain the convergence of the specific enthalpies of denaturation. [Recently, Lee (1991), Doig and Williams (1991), and Baldwin and Muller (1992) have also addressed the issues raised by Murphy et al. (1990).]

Methods

All water-accessible surface areas were calculated as previously described (Livingstone et al., 1991), using group radii from Richards (1974). All carbon atoms are considered to be nonpolar; all nitrogen and oxygen atoms are considered to be polar. To simplify the notation, the constant-pressure heat capacity change is designated ΔC°, and the symbol $C^\circ_{x,p}$ indicates the contribution to ΔC° for process x (transfer, folding) from changes in water-accessible *polar* surface.

Results and Discussion

Choice of Proteins for Analysis

To eliminate the possibility that the apparent discrepancy between the thermodynamic behaviors of liquid hydrocarbons and proteins results in part from the choice of proteins in the data set, we first examined whether the proteins used to test the applicability of the liquid hydrocarbon model at the level of ΔC°_{fold} (Livingstone et al., 1991) yield the linear relationship between the specific quantities Δh°_{fold} and Δc°_{fold} noted by Murphy et al. (1990). The proteins shown in Table I were selected by Livingstone et al. (1991) because both Δc°_{fold} and the high-resolution crystal structure of the native state are known. With the exception of fragment K4 of plasminogen, the set of proteins analyzed by Murphy et al. (1990) is entirely contained in our set. Figure 1 plots Δh°_{fold} at 298 K for the proteins of Table I versus Δc°_{fold}. Only ribonuclease T_1 deviates markedly from a linear relationship. A linear fit of the protein data excluding ribonuclease T_1 has a slope of $(298 - T_H^*) = -59 \pm 9$ K and an intercept of -9.6 ± 1.2 cal g^{-1}. Hence, for this set of proteins, Δh°_{fold} of folding has a constant value of -9.6 ± 1.2 cal g^{-1} at $T_H^* = 357 \pm 9$ K. Does this property of Δh°_{fold} result from the thermodynamic consequences of removal of polar and nonpolar surface from water upon folding?

Contribution of Changes in Water-Accessible Polar Surface Area to Heat Capacity Changes of Transfers of Organic Amides and Folding of Proteins

Organic Amide Transfer Process (Aqueous → Pure Liquid). Since changes in the water accessibility of the carbonyl oxygen and the nitrogen of the amide group in the peptide backbone comprise the majority (~70%) of the change in polar surface area upon folding a globular protein (Lee & Richards, 1971; Shrake & Rupley, 1973), we have examined

Table I: Values of ΔA_{np}, ΔA_p, ΔC°_{fold}, and ΔC°_{calc} for Protein Folding

protein	M_r (g mol^{-1})	ΔA_{np} (Å^2)	ΔA_p (Å^2)	$-\Delta C^\circ_{fold}$ [a] (cal mol^{-1} K^{-1})	$-\Delta C^\circ_{calc}$ [b] (cal mol^{-1} K^{-1})
trypsin inhibitor	6520	2640	1500	720 ± 110[c]	635 ± 120
parvalbumin B	11500	5485	3550	1100 ± 120	1265 ± 260
ribonuclease A	13700	5815	4475	1230 ± 180	1235 ± 290
lysozyme, egg white	14300	6870	4360	1540 ± 190	1590 ± 320
ribonuclease T_1	10900	5165	3040	1650 ± 200[d]	1230 ± 240
ferricytochrome	12300	5540	3300	1730 ± 260	1310 ± 260
staphylococcal nuclease	16800	7880	5860	1820 ± 230[e]	1700 ± 390
metmyoglobin	17800	9710	5245	2770 ± 420	2370 ± 440
β-trypsin	23200	11830	8495	2850 ± 440	2600 ± 580
papain	23400	12755	7325	2920 ± 450	3060 ± 590
α-chymotrypsin	25200	14770	7835	3020 ± 500	3630 ± 670
chymotrypsinogen	25700	14550	7275	3800 ± 260[f]	3640 ± 650
carbonic anhydrase	28400	15760	9210	3820 ± 150	3750 ± 730
pepsinogen	40000	23730	11645	6090 ± 300	5960 ± 1060

[a] All heat capacity data are from Privalov and Gill (1988) unless otherwise noted. [b] $\Delta C^\circ_{calc} = -(0.32 \pm 0.04)\Delta A_{np} + (0.14 \pm 0.04)\Delta A_p$. [c] Privalov (1979). [d] Pace and Laurents (1989). [e] Griko et al. (1988). [f] Hawley (1971).

Table II: Values of A_{np}, A_p, ΔH°_{tr}, ΔC°_{tr}, and ΔC°_{calc} for Transfers of Amides at 298 K

compound	A_{np} [a] (Å^2)	A_p [a] (Å^2)	ΔH°_{tr} [b] (kcal mol^{-1})	ΔC°_{tr} [b] (cal mol^{-1} K^{-1})	ΔC°_{calc} [c] (cal mol^{-1} K^{-1})
formamide	39.9	122	−0.485 ± 0.002[d]	6.12 ± 0.24[d]	7.10 ± 4.9
N-methylformamide	128	69.9	1.693 ± 0.001[d]	−9.61 ± 0.48[d]	−22.2 ± 3.8
N-methylacetamide	182	54.2	3.129 ± 0.005[d]	−25.48 ± 0.72[d]	−37.9 ± 4.2
N-ethylacetamide	217	50.8	3.700 ± 0.005	−38.96 ± 0.96	−47.2 ± 4.8
N-propylacetamide	247	50.8	3.767 ± 0.005	−54.97 ± 0.96	−54.6 ± 5.3
N-isopropylacetamide	250	45.5	4.120 ± 0.005	−54.97 ± 0.96	−56.0 ± 5.3
N-butylacetamide	277	48.8	3.518 ± 0.007[d]	−66.92 ± 0.34[d]	−62.4 ± 5.9
N-methylpropanamide	216	49.0	3.554 ± 0.005	−37.05 ± 0.72	−47.1 ± 4.7
N-methylbutanamide	245	49.0	3.829 ± 0.005	−54.25 ± 0.96	−54.4 ± 5.3
N,2-dimethylpropanamide	246	43.2	3.774 ± 0.005	−53.06 ± 1.20	−55.4 ± 5.2
N-methylpentanamide	275	47.0	3.592 ± 0.005[d]	−68.26 ± 0.96[d]	−62.2 ± 5.8

[a] Water-accessible surface areas calculated as described by Livingstone et al. (1991). [b] Data are from Konicek and Wadsö (1971) unless noted otherwise. [c] $\Delta C^\circ_{calc} = -(0.25 \pm 0.02)\Delta A_{np} + (0.14 \pm 0.04)\Delta A_p$. [d] Data from Sköld et al. (1976).

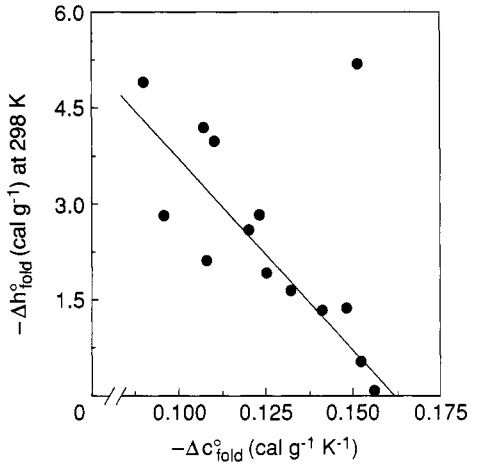

FIGURE 1: Specific enthalpy of folding (Δh°_{fold}) at 298 K (Table III) as a function of specific heat capacity of folding (Δc°_{fold}) (Table I). The solid line corresponds to the least-squares fit of the data and conforms to the equation $\Delta h^\circ_{fold} = -(9.6 \pm 1.2 \text{ cal g}^{-1}) - (59 \pm 9 \text{ K})\Delta c^\circ_{fold}$.

the relationship between the change in polar surface area and the observed heat capacity change for the transfer of various organic amides from water to the pure liquid phase. Table II shows all available pairs of calorimetrically determined values of ΔC°_{tr} and ΔH°_{tr} for the transfer of amides from the infinitely dilute aqueous solution to the pure liquid phase, along with the calculated values of the water-accessible nonpolar (A_{np}) and polar (A_p) surface areas of each compound. Analysis of analogous transfer data for nonpolar liquid hydrocarbons indicates that $\Delta C^\circ_{tr,np}$ is directly proportional to A_{np} (cf. eq 3) (Livingstone et al., 1991). We therefore asked whether generalization of this relationship to include a contribution from polar surface area (A_p) would fit the organic amide transfer data. We assume that contributions to ΔC°_{tr} from polar and nonpolar surface are additive and that each is proportional to the relevant water-accessible surface area. The least-squares fitting program NONLIN (Johnson & Frasier, 1985) was used to fit ΔC°_{tr} as a function of A_{np} and A_p:

$$\Delta C^\circ_{tr} = \Delta C^\circ_{tr,np} + \Delta C^\circ_{tr,p} = a_{tr}A_{np} + b_{tr}A_p \qquad (9)$$

The values of a_{tr} and b_{tr} and their 67% confidence intervals from this fit are $a_{tr} = -0.25$ (−0.27, −0.23) cal mol^{-1} K^{-1} Å^{-2} and $b_{tr} = 0.14$ (0.11, 0.18) cal mol^{-1} K^{-1} Å^{-2}. Calculated heat capacity changes, $\Delta C^\circ_{calc} \equiv -0.25A_{np} + 0.14A_p$, for individual organic amides are plotted versus the experimental ΔC°_{tr} in Figure 2. The solid line in Figure 2 has a unit slope and a zero intercept, representing perfect agreement between calculated and observed values. Though the data may exhibit some systematic deviations from this line, we consider the fit to be reasonable, given the assumptions of the analysis and the fact that the data set is limited in the amount of variability in A_p. In addition, the coefficient $a_{tr} \equiv (\partial \Delta C^\circ_{tr}/\partial A_{np})_{A_p} = -0.25 \pm 0.04$ cal mol^{-1} K^{-1} Å^{-2} from this fit agrees within two standard deviations with that obtained for compounds composed only of nonpolar surface [$a_{tr} = -(0.32 \pm 0.08)$ cal mol^{-1} K^{-1} Å^{-2}; Livingstone et al., 1991]. For organic amides, the above analysis indicates that the contribution to ΔC°_{tr} from burial of polar surface (per Å^2) is approximately 60% as large and has the opposite sign as that from burial of nonpolar surface.

Protein Folding. Calculations of the amounts of water-accessible polar and nonpolar surface area buried upon folding of single-domain globular proteins (cf. Table I) indicate that

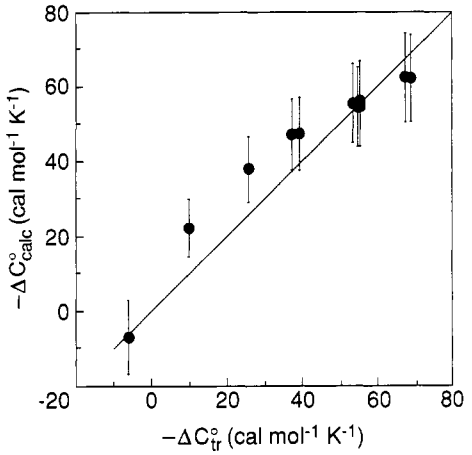

FIGURE 2: Calculated heat capacity change (ΔC°_{calc}) from Table II for the process of transferring amides from the infinitely dilute solution to the pure liquid state plotted versus the observed standard heat capacity change (ΔC°_{tr}). Error bars represent two standard deviations. Perfect agreement between the observed and calculated values is represented by the solid line, which has a slope of 1 and a y-intercept of 0.

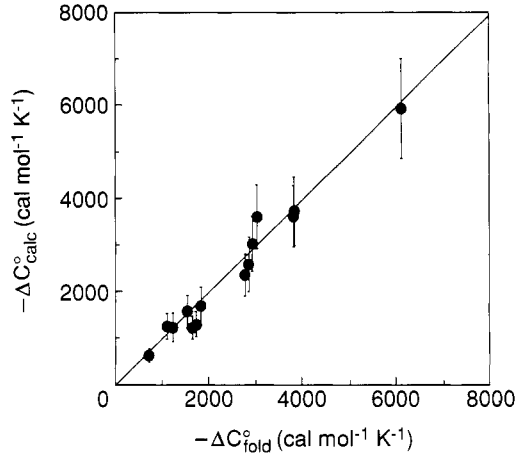

FIGURE 3: Comparison of the observed standard heat capacity change for the process of folding (ΔC°_{fold}) the proteins shown in Table I with the heat capacity change calculated from the equation $\Delta C^\circ_{calc} = -0.32\Delta A_{np} + 0.14\Delta A_p$ (●). Error bars represent one standard deviation. The solid line represents perfect agreement between calculated and observed values.

~40% of the total change in water-accessible surface area (ΔA_T) is polar (ΔA_p) whereas ~60% is nonpolar (ΔA_{np}). If the organic amide transfer data serve as a useful model for heat capacity effects in protein folding, then the contributions to ΔC°_{fold} from burial of nonpolar and polar surface ($\Delta C^\circ_{fold,np}$ and $\Delta C^\circ_{fold,p}$) should fit the following generalization of eq 9:

$$\Delta C^\circ_{fold} = \Delta C^\circ_{fold,np} + \Delta C^\circ_{fold,p} + \Delta C^\circ_{other} \quad (10)$$

where $\Delta C^\circ_{fold,np} = a_{fold}\Delta A_{np}$ and $\Delta C^\circ_{fold,p} = b_{fold}\Delta A_p$ and where a_{fold} and b_{fold} should be the same within error as the corresponding coefficients evaluated from transfer data (eqs 3 and 9). In order to evaluate a_{fold}, b_{fold}, and ΔC°_{other}, values of ΔC°_{fold}, ΔA_{np}, and ΔA_p for all single-domain proteins for which calorimetric and structural data are available (compiled in Table I) were fit to eq 10. Values of b_{fold} and ΔC°_{other} were either allowed to float along with a_{fold} or were fixed at zero; the relative quality of the various fits was determined by comparing the reduced chi-squared term (χ^2) for each fit. If the data are assumed to follow a normal Gaussian distribution, then the closer χ^2 is to one, the better the fit approximates the data (Snedecor & Cochran, 1980; Press et al., 1986). Using this criterion, the worst fit of these data corresponds to setting b_{fold} equal to zero for which $\chi^2 = 1.369$. The best fit of the data occurs when ΔC°_{other} is set to zero, for which $\chi^2 = 1.033$. The values of a_{fold} and b_{fold} from this best fit and their corresponding 67% confidence intervals are $a_{fold} = -0.33$ (−0.41, −0.26) cal mol^{-1} K^{-1} Å^{-2} and $b_{fold} = 0.16$ (0.034, 0.28) cal mol^{-1} K^{-1} Å^{-2}. These values agree within error with those obtained from the analysis of model amide and hydrocarbon transfer data. Moreover, they are consistent with our previous analysis of ΔC°_{fold} as a function of ΔA_{np} (cf. eq 4), because ΔA_{np} and ΔA_p are proportional to one another and to ΔA_T for the proteins in the data set in Table I: $\Delta A_p = (0.59 \pm 0.08)\Delta A_{np} = (0.38 \pm 0.03)\Delta A_T$. Therefore, for these proteins eq 10 may be rewritten

$$\Delta C^\circ_{fold} = -0.33\Delta A_{np} + 0.16\Delta A_p$$

$$\Delta C^\circ_{fold} \approx [-0.33 + 0.59(0.16)]\Delta A_{np} \approx -0.24\Delta A_{np} \quad (11)$$

Equation 11 agrees with our previous observation that $\Delta C^\circ_{fold} = -(0.25 \pm 0.03)\Delta A_{np}$ (Livingstone et al., 1991). The burial of nonpolar surface is the dominant contributor to ΔC°_{fold}. The relationship between ΔA_p and ΔA_{np} for protein folding allows the entire effect to be parameterized in terms of ΔA_{np}, with a proportionality constant which is the same within error as that of the liquid hydrocarbon data.

Figure 3 demonstrates the quality of agreement between the liquid hydrocarbon/amide transfer data and protein folding data by plotting the predicted heat capacity change (ΔC°_{calc}) of protein folding versus the observed heat capacity change (ΔC°_{fold}), using the value of a_{tr} from the liquid hydrocarbon data and b_{tr} from the organic amide transfer data[1] (uncertainties are ±1 SD):

$$\Delta C^\circ_{calc} = -(0.32 \pm 0.04)\Delta A_{np} + (0.14 \pm 0.04)\Delta A_p \text{ cal mol}^{-1}\text{ K}^{-1} \quad (12)$$

We recognize that the uncertainty in b_{fold} is sufficiently large that b_{fold} may be negligibly small and that the aforementioned proportionality between ΔA_p and ΔA_{np} of folding for our data set reduces the number of independent variables and hence introduces a source of ambiguity into the analysis. However, given the agreement shown in Figure 3, we propose that (1) changes in water-accessible *polar* surface contribute to ΔC°_{fold} and (2) transfers of model organic amides and hydrocarbons from water to the pure liquid phase provide useful quantitative models of the heat capacity change in protein folding. This refinement appears warranted since the original liquid hydrocarbon model does not explain the observed convergence of the specific enthalpies of folding at T_H*, as noted by Baldwin (1986). (However, see Added in Proof.) In what follows, we propose a means of estimating both surface-

[1] To examine whether any systematic difference exists between the values of a_{tr} and b_{tr} obtained from the independent analyses of the amide data set in Table II and the hydrocarbon data set in Livingstone et al. (1991), the two data sets were combined and fit to eq 10 using NONLIN. The values of a_{tr}, b_{tr}, and ΔC_{other} from this fit are $a_{tr} = -0.28$ (−0.22, −0.34) cal mol^{-1} K^{-1} Å^{-2}; $b_{tr} = 0.09$ (0.05, 0.22) cal mol^{-1} K^{-1} Å^{-2}; $\Delta C_{other} = 8.8$ (−11, 28) cal mol^{-1} K^{-1}. These values of a_{tr} and b_{tr} agree within uncertainty with those from the independent analyses and with those from the analysis of the protein data set. There is no systematic difference between a_{tr} for the two model compound data sets. In addition, ΔC_{other} is zero within uncertainty, consistent with our simplifying assumption that the contributions to ΔC°_{tr} from changes in water-accessible nonpolar and polar surface are additive and that these account for the entire ΔC°_{tr}. Equation 12 is written using the values from the independent analyses of the hydrocarbons (a_{tr}) and amides (b_{tr}) rather than the above parameters from the composite fit, because the two data sets have significantly different associated uncertainties. However, the conclusions of this analysis are insensitive to the exact values of these coefficients.

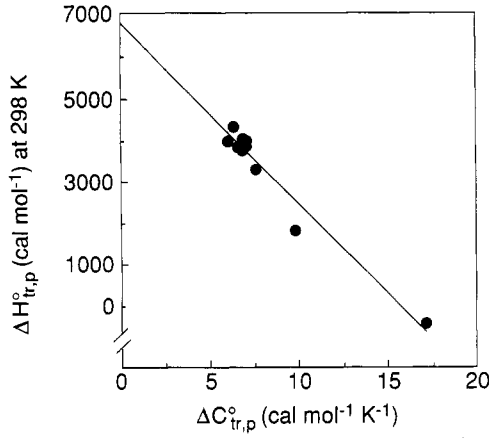

FIGURE 4: Polar contribution ($\Delta H^{\circ}_{tr,p}$) to the observed enthalpy of transfer of amides from the infinitely dilute solution to the pure liquid phase as a function of $\Delta C^{\circ}_{tr,p}$, where $\Delta C^{\circ}_{tr,p} = 0.14A_p$.

area-dependent and intrinsic *polar* contributions to the *enthalpy* of folding from organic amide transfer data.

Analysis of the Contributions of Polar Surface Area to the Enthalpy of Transfer of Liquid Amides

In the liquid hydrocarbon model, the nonpolar contribution to the enthalpy of transfer ($\Delta H^{\circ}_{tr,np}$) is estimated by eq 6, where $\Delta C^{\circ}_{tr,np} = -0.32\Delta A_{np}$ (Livingstone et al., 1991). Does an analogous expression exist for estimating the *polar* contribution to the enthalpy of transfer? If so, then the polar contribution ($\Delta H^{\circ}_{tr,p}$) to ΔH°_{tr} should be linearly related to the polar contribution ($\Delta C^{\circ}_{tr,p}$) to ΔC°_{tr} (cf. eq 9). Values of $\Delta H^{\circ}_{tr,p}$ for the transfer of organic amides were evaluated from experimental values of ΔH°_{tr} at 298 K:

$$\Delta H^{\circ}_{tr,p}(298) = \Delta H^{\circ}_{tr}(298) - \Delta H^{\circ}_{tr,np}(298) \quad (13)$$

where $\Delta H^{\circ}_{tr,np}(298)$ is obtained from eq 6. A plot of $\Delta H^{\circ}_{tr,p}(298)$ versus $\Delta C^{\circ}_{tr,p}$, where $\Delta C^{\circ}_{tr,p} = 0.14A_p$ for these compounds (cf. eq 9) is shown in Figure 4 and yields the following linear least-squares fit:

$$\Delta H^{\circ}_{tr,p}(298) = (298 - 735 \pm 1)\Delta C^{\circ}_{tr,p} + 6.820 \pm 0.001 \text{ kcal mol}^{-1} \quad (14)$$

(The small associated uncertainties in eq 14 are omitted subsequently.) Equation 14 demonstrates that the polar contribution to the enthalpy of transfer of the organic amides is a linear function of the polar contribution to the heat capacity of transfer, with a characteristic temperature $T_H^* = 735$ K, at which temperature $\Delta H^{\circ}_{tr,p}(735) = 6.82$ kcal mol^{-1} for all organic amides. This also is the value of $\Delta H^{\circ}_{tr,p}$ for organic amides in the limit of zero polar surface area (Lee, 1991) and hence corresponds to an intrinsic polar contribution to the enthalpy of transfer of the amide group ($\Delta H^{int}_{tr,p}$) which is independent of polar surface area. Consequently eq 14 may be rewritten as

$$\Delta H^{\circ}_{tr,p}(298) = (298 - 735)\Delta C^{\circ}_{tr,p} + \Delta H^{int}_{tr,p} \quad (15)$$

Use of Transfer Data To Estimate The Molar Enthalpy ΔH°_{fold} of Protein Folding

We propose that eq 15 may be applied to protein folding (and to other processes) to estimate the thermodynamic consequences of changes in the water-accessible polar surface area of amide groups at 298 K:

$$\Delta H^{\circ}_{fold,p}(298) = 0.14(298 - 735)\Delta A_p + n\Delta H^{int}_{fold,p} \quad (16)$$

where n is the number of amide groups buried and $n\Delta H^{int}_{fold,p}$ is the surface-area-independent (i.e., intrinsic) and temperature-independent contribution of an amide group to the polar part of the enthalpy of folding. More generally, at any temperature T

$$\Delta H^{\circ}_{fold,p}(T) = 0.14(T - 735)\Delta A_p + n\Delta H^{int}_{fold,p} \quad (17)$$

Since a surface-area-independent quantity such as $\Delta H^{int}_{fold,p}$ is likely to be process dependent (i.e., $\Delta H^{int}_{tr,p} \neq \Delta H^{int}_{fold,p}$), we propose that the model organic amide and hydrocarbon transfer data be utilized to estimate the contributions to ΔH°_{fold} which depend upon temperature and surface area and that protein folding data be used to estimate $\Delta H^{int}_{fold,p}$. [This approach is similar to that taken by Ooi and Oobatake (1988), although not explicitly stated by them.] If the liquid amide/hydrocarbon model of protein unfolding is appropriate, then the net enthalpy of folding at any temperature T is given by

$$\Delta H^{\circ}_{fold}(T) = (T - 295)\Delta C^{\circ}_{fold,np} + (T - 735)\Delta C^{\circ}_{fold,p} + n\Delta H^{int}_{fold,p}$$

$$\Delta H^{\circ}_{fold}(T) = -0.32(T - 295)\Delta A_{np} + 0.14(T - 735)\Delta A_p + n\Delta H^{int}_{fold,p} \quad (18)$$

Contributions to $\Delta H^{\circ}_{fold}(298)$ from the burial of nonpolar surface [$-0.32(298 - 295)\Delta A_{np}$] and polar surface [$0.14(298 - 735)\Delta A_p$] are collected in Table III. At 298 K the contribution from burial of polar surface is large and negative, and the sum of the polar and nonpolar contributions is much larger in magnitude than the observed value of ΔH°_{fold}. Subtracting them from ΔH°_{fold} yields a large positive residual enthalpy change $n\Delta H^{int}_{fold,p}$. Lee and Richards (1971) determined accessibilities of the amide nitrogen and the carbonyl oxygen in the folded state of several representative proteins. Their results indicate that nearly all of the amide backbone is buried upon folding. In this case, $n\Delta H^{int}_{fold,p}$ corresponds to an enthalpy increase $\Delta H^{int}_{fold,p} \approx 2$ kcal/mol of backbone amide groups buried upon folding (cf. Table III).

One possible interpretation of $\Delta H^{int}_{fold,p}$ is provided by studies of the number of hydrogen bonds existing in native proteins. Work by Chothia (1975) and Privalov (1979) indicates that each amino acid residue makes an average of only 0.75–1 hydrogen bond in the native state. For the amide backbone, Baker and Hubbard (1984) find that approximately 12% of the carbonyl oxygens and amide nitrogens do not form any hydrogen bonds and that 53% of the carbonyl oxygens form only one hydrogen bond. Given the lack of steric constraint of the denatured state relative to the native state and the presence of water solvating the chain, it is plausible that the carbonyl oxygen and amide nitrogen participate in more hydrogen bonds in the denatured state. It is not known how the free energy of formation of these compares with that of the water–water hydrogen bonds they would replace. We propose that $\Delta H^{int}_{fold,p}$ represents the difference in quality and quantity of hydrogen bonds made in the native and denatured states. Ideally, one would like to model the hydrogen-bonding contribution to protein folding using accurate thermodynamic measurements made on well-designed model systems. Conclusions based on experiments designed by Klotz and coworkers (Klotz & Franzen, 1962; Klotz & Farnham, 1968; Kreshek & Klotz, 1969) to measure the free energy of forming a hydrogen bond between *N*-methylacetamides in various solvents are ambiguous and have been subject to reinterpretation (Roseman, 1988). Although the question of whether hydrogen bonds contribute to protein stability and how much they contribute is controversial, our analysis indicates that the enthalpic contribution from hydrogen bonding may favor

Table III: Values of ΔH°_{fold} and Calculated Contributions from Burial of Nonpolar and Polar Surface at 298 K[a]

		contributions to ΔH°_{fold}			
			polar		
protein	ΔH°_{fold}[b]	nonpolar[c]	surface area dependent[d]	intrinsic (surface area independent)[e]	$\Delta H^{int}_{fold,p}$
trypsin inhibitor	−26[f]	−2.54	−91.8	68.3	1.2
parvalbumin B	−32.5	−5.26	−214	187	1.7
ribonuclease A	−67.0	−5.58	−274	212	1.7
lysozyme, egg white	−60.0	−6.59	−267	213	1.7
ferricytochrome	−16.7	−5.32	−202	191	1.8
staphylococcal nuclease	−35.8[g]	−7.57	−358	330	2.3
metmyoglobin	−1.49	−9.32	−321	330	2.2
β-trypsin	−65.8	−11.4	−519	465	2.1
papain	−45.2	−12.2	−448	415	2.0
α-chymotrypsin	−66.0	−14.2	−479	427	1.8
chymotrypsinogen	−35.6[h]	−14.0	−445	423	1.7
carbonic anhydrase	−48.2	−15.1	−563	530	2.1
pepsinogen	−22.7[i]	−22.8	−712	712	1.9

[a] All entries are in units of kcal mol^{-1}. [b] Values of ΔH°_{fold} at 298 K are from Privalov and Gill (1988) unless otherwise noted. [c] Nonpolar contribution = $-0.32(298 - 295)\Delta A_{np}$. [d] Polar-surface-area dependent contribution = $0.14(298 - 735)\Delta A_{p}$. [e] Intrinsic polar contribution = $n\Delta H^{int}_{fold,p}$. [f] Privalov (1979). [g] Griko et al. (1988) [h] Hawley (1971). [i] Mateo and Privalov (1981).

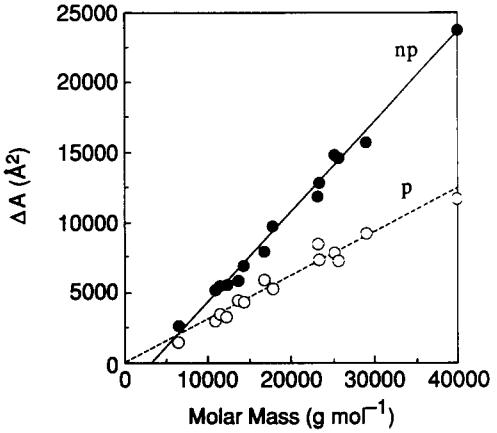

FIGURE 5: Change in the water-accessible nonpolar surface area ΔA_{np} (●) and the change in the water-accessible polar surface area ΔA_{p} (○) accompanying the folding of the proteins in Table I as a function of molar mass. Lines represent least-squares fits (cf. eqs 19–20 in the text).

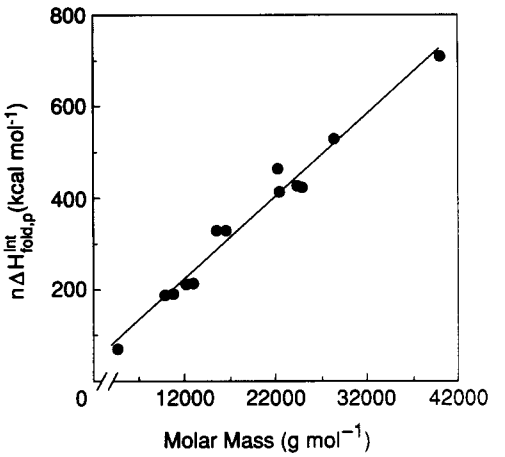

FIGURE 6: Contribution to ΔH°_{fold} from intrinsic factors upon burial of amide groups ($n\Delta H^{int}_{fold,p}$) (Table III) as a function of molar mass. Line represents linear least-squares fit (cf. eq 21 in the text).

protein unfolding, as suggested by Dill (1990).

Convergence of Specific Enthalpies of Folding (Δh°_{fold})

The observation that the specific enthalpy of folding Δh°_{fold} has a constant value of ~−9.6 cal g^{-1} at 357 K for this protein data set led us to examine the dependences of ΔA_{np}, ΔA_{p}, and $n\Delta H^{int}_{fold,p}$ on molar mass (M_r). For the set of proteins in Table I, ΔA_{np}, ΔA_{p}, and $n\Delta H^{int}_{fold,p}$ (cf. Figures 5 and 6) are all linear functions of molar mass in the range of interest (6500–40 000 g mol^{-1}):

$$\Delta A_{np} = (0.636 \pm 0.018)M_r - (2120 \pm 380)\ \text{Å}^2 \quad (19)$$

$$\Delta A_{p} = (0.309 \pm 0.018)M_r - (40 \pm 380)\ \text{Å}^2 \quad (20)$$

$$n\Delta H^{int}_{fold,p} = (0.019 \pm 0.001)M_r - (36 \pm 19)\ \text{kcal mol}^{-1} \quad (21)$$

The sum of eqs 19–20 for the change in total surface area (ΔA_T) upon folding differs in functional form from that reported by Chothia (1975) and Miller et al. (1987). However, we find that their data can be adequately fit as a linear function of molar mass with coefficients which agree with the sum of eqs 19–20 (data not shown). In the limit of large molar mass, eq 21 is consistent with the observation in Table III that $\Delta H^{int}_{fold,p} \approx 2$ kcal/mol of buried backbone amide groups.

From eqs 18–21, we obtain a prediction for the dependence of the specific enthalpy of folding on molar mass and temperature, quantitatively applicable to proteins for which eqs 12 and 19–21 are accurate.

$$\Delta h^{\circ}_{fold}(T) = \Delta H^{\circ}_{fold}(T)/M_r = -0.161T + 47.24 + 672.8T/M_r - 232010/M_r\ \text{cal g}^{-1} \quad (22)$$

where round-off is avoided to preserve the significance of the calculated Δh°_{fold}. Equation 22 predicts that $\Delta h^{\circ}_{fold}(T)$ will be a constant at the temperature where the inverse molar mass terms cancel (i.e., where $672.8T = 232010$ or $T = 345$ K). At 345 K, $\Delta h^{\circ}_{fold} = \Delta H^{\circ}_{fold}/M_r = -8.3$ cal g^{-1}. This result, which is derived from transfer data for liquid hydrocarbons and amides as well as from protein folding data, agrees reasonably well with that obtained by extrapolation of the experimental data: $\Delta H^{\circ}_{fold}/M_r = -9.6 \pm 1.2$ cal g^{-1} at 357 ± 9 K.

Predictive Value of Relationships Based on Model Compound Transfer Data

Enthalpy and Heat Capacity. Table IV provides an independent test of the validity of this liquid hydrocarbon/amide model by comparing values of ΔC°_{fold} and ΔH°_{fold} calculated using eqs 12, 19, 20, and 22 with those measured calorimetrically for proteins not in the data set in Table I. These proteins are not included in Table I because they lack a high-resolution crystal structure. Proteins chosen for inclusion

Table IV: Comparison of Calorimetric Values of $\Delta H^{\circ}_{fold}(T)$ and ΔC°_{fold} with Those Calculated Using the Liquid Hydrocarbon/Amide Transfer Model

protein	M_r (g mol^{-1})	T (K)	$-\Delta H^{\circ}_{fold}$ [a] (kcal mol^{-1})	$-\Delta H^{\circ}_{calc}$ (kcal mol^{-1})	$-\Delta C^{\circ}_{fold}$ [a] (cal mol^{-1} K^{-1})	$-\Delta C^{\circ}_{calc}$ (cal mol^{-1} K^{-1})
Ac-Y(AEAAKA)$_8$F-NH$_2$ [b]	4682	314	43.2	36.3	ND	77.6
GH1 tryptic fragment of histone H1 [c]	8590	332.65	44	62	1100	710
GH5 tryptic fragment of histone H5 [c]	8903	336.15	51	67	1060	760
fragment K4 of plasminogen [d]	9640	316	51 ± 3	54	955 ± 170	875
hydrophilic fragment of cytochrome b_5 [e]	10933	378	130	127	1460 ± 220	1080
λ repressor [f]						
N-terminal domain	10125	322	63	62	ND	950
C-terminal domain	15977	346	140	135	ND	1890

[a] Error is ±1 SD; ND, not determined. [b] Scholtz et al. (1991). [c] Tiktopulo et al. (1982). [d] Novokhatny et al. (1984). [e] Pfeil and Bendzko (1980). [f] Pabo et al. (1979).

Table V: Representative Comparisons of Protein Stability (ΔG°_{fold}) with That Predicted from Molar Mass (M_r) and T_m [a]

		noncalorimetric			calorimetric		predicted
	M_r (g mol^{-1})	pH	T_m (K)	$-\Delta G^{\circ}_{fold}(T)$ (kcal mol^{-1})	T_m (K)	$-\Delta G^{\circ}_{fold}(T)$ (kcal mol^{-1})	$-\Delta G^{\circ}_{calc}$ (kcal mol^{-1})
RNase A	13690	2.8	315.75 [b]	4.33 (298.05) [b]	315.15 [c]	4.42 (298.05) [c]	3.0
lysozyme, egg white	14320	7.0	ND	14.2 (298.15) [d]	353.15 [e]	14.5 ± 0.8 (298.15) [e]	13.2
staphylococcal nuclease	16800	7.0	326.45 [f]	6.1 (293.15) [f]	324.85 [g]	5.33 (293.15) [g]	6.4
lysozyme, T4 phage	18700	3.0	329.80 [h]	2.35 (320.00) [h]	326.71 [i]	2.27 (320.00) [i]	2.2
α-chymotrypsin	25200	4.0	ND	10.8 (298.15) [j]	330.25 [k]	11.6 ± 0.5 (298.15) [e]	10.1
				11.7 (298.15) [j]			
pepsinogen	40000	6.5	ND	6.5 (298.15) [l]	337.65 [m]	15.7 (298.15) [m]	20.2
				7.6 (298.15) [l]			
				11.7 (298.15) [l]			

[a] Abbreviation: ND, not determined. Equations 2, 12, 19, 20, and 22 and the calorimetrically determined value of T_m were used to calculate ΔG_{calc}. [b] Pace and Laurents (1989). [c] Privalov et al. (1973). [d] Salahuddin and Tanford (1970). [e] Pfeil and Privalov (1976). [f] Shortle et al. (1988). [g] Griko et al. (1988). [h] Hawkes et al. (1984). [i] Connelly et al. (1991). [j] Pfeil (1986). [k] Privalov and Khechinashvili (1974). [l] Ahmad and McPhie (1978). [m] Privalov et al. (1981).

in Table IV are ones for which the ratio of the calorimetric ΔH°_{fold} to the van't Hoff ΔH°_{fold} is approximately 1, indicating that folding of these proteins is a two-state transition. (We are not aware of other single-domain proteins which satisfy the above criteria for inclusion in Table IV.) All values of ΔC°_{calc} and ΔH°_{calc} in Table IV are predicted solely from molar mass. For a protein with a M_r of 20000 g mol^{-1}, propagation of the uncertainties in eqs 12 and 19–20 gives a relative uncertainty of ±24% in ΔC°_{calc}. For comparison, the typical error on a calorimetrically measured ΔC°_{fold} ranges from 5% to 20% (Becktel & Schellman, 1987; Pace & Laurents, 1989). The relative uncertainties associated with the nonpolar, polar, and intrinsic contributions to ΔH°_{fold} are each ~20%, but propagation of these uncertainties yields a relative uncertainty in ΔH°_{calc} (~85%) which far exceeds the experimental uncertainty in ΔH°_{fold} (~5%) (Becktel & Schellman, 1987; Pace & Laurents, 1989). Using the average experimental uncertainty as a criterion, ΔH°_{calc} for the last four proteins in Table IV appears to agree with ΔH°_{fold}, whereas the values of ΔH°_{calc} for the tryptic fragments of histones H1 and H5 do not, and the value of ΔH°_{calc} for the peptide appears marginally acceptable. The successful use of M_r to obtain quantitative estimates of ΔC°_{fold} and ΔH°_{fold} ultimately depends on whether the protein of interest obeys eqs 12 and 19–21. Consequently, the lack of agreement between prediction and experiment for the histone fragments may indicate that either the surface area–thermodynamic relationships and/or the surface area–molar mass relationships do not describe these proteins.

Free Energy. Given an accurate estimate of H°_{fold} and ΔC°_{fold}, a single determination of T_m for a single-domain protein of known molecular mass allows one to predict ΔG°_{fold} as a function of temperature for the solution conditions of that measurement (cf. eq 2). For purposes of illustration, Table V compares representative values of ΔG°_{fold} for single-domain globular proteins whose stability has been determined *both* by titration with chemical denaturants and by calorimetry under the same solution conditions with those calculated using eqs 2, 12, 19, 20, and 22 and the calorimetric value of T_m. The proteins in Table V span a wide range of molar masses and stabilities as well as the range of values of $n\Delta H^{int}_{fold,p}$ (cf. Table III). Since measurements of ΔG°_{fold} as a function of denaturant concentration and subsequent extrapolations to determine ΔG°_{fold} in the absence of denaturant have an associated uncertainty of approximately ±15% (Pace, 1975), we conclude that, for the proteins in Table V, this method provides acceptable estimates of protein stability. A detailed comparison of values of ΔG°_{fold} determined using calorimetry, denaturant titration, and/or van't Hoff analysis with those calculated by the present method is beyond the scope of this paper. A large data set for such a comparison is reviewed by Pfeil (1986). Deviation from the predicted thermodynamic behavior which may be observed for other proteins should in principle be interpretable in terms of deviations from the average structural (eqs 19–21) and/or thermodynamic (eq 12) properties of the proteins considered here. In particular, we find that calculated values of ΔG°_{fold} for both heme proteins in our data set differ by ~5 kcal mol^{-1} from the calorimetric values. These discrepancies appear to result from inaccurate estimates of ΔH°_{fold} using eq 22. In addition, given that the model's usefulness is limited to proteins which obey the "average" behavior of the data set analyzed here, it is almost certainly not applicable to analyses of the thermodynamic differences between wild-type and mutant proteins. The model and analyses developed here seek to identify and predict the major contributions to the thermodynamics of protein folding from removal of polar and nonpolar surface from water.

Conclusion

Although the change in water-accessible nonpolar surface area makes the dominant contribution to ΔC°_{fold}, we propose

that changes in the water-accessibility of the peptide backbone also contribute to this fundamental determinant of the large temperature dependences of the enthalpy and entropy of folding. This refinement of our previous analysis (Spolar et al., 1989; Livingstone et al., 1991) allows the following conclusions to be drawn: (1) transfers of *both* polar and nonpolar model compounds from water to the pure liquid phase (the liquid hydrocarbon/amide model) appropriately model protein folding at the level of ΔC°_{fold}; (2) specific enthalpies of folding (Δh°_{fold}) are predicted by the liquid hydrocarbon/amide model to converge to a value of −8.3 cal g^{-1} at T_H^* = 345 K, whereas the experimental values are Δh°_{fold} = −9.6 ± 1.2 cal g^{-1} at T_H^* = 357 ± 9 K. In addition, for proteins whose surface area–heat capacity and surface area–molar mass relationships are described by the average behavior of the protein data set considered here: (1) values of ΔC°_{fold} and ΔH°_{fold} may be estimated from molar mass alone and (2) from one accurate determination of T_m and the molar mass of the protein, application of the liquid hydrocarbon/amide model provides a reasonable estimate of ΔG°_{fold} as a function of temperature for the experimental conditions of interest.

Added in Proof

Baldwin and Muller (1992) have derived a thermodynamic relationship between ΔH°_{fold} and ΔS°_{fold} at the convergence temperature T_s^* where the specific entropies of folding of a set of proteins are the same. This derivation, which is independent of the liquid hydrocarbon model, indicates that the specific enthalpies of folding of this set of proteins should converge at a slightly lower temperature.

Acknowledgments

We thank H. Guttman and Dr. V. Licata for help with computing and statistical analysis and the referees for their constructive criticisms of the manuscript. We thank Drs. Baldwin and Muller for their comments on this paper and for sending us their manuscript in advance of publication. We appreciate Sheila Aiello's help in preparation of the manuscript.

Registry No. trypsin inhibitor, 9035-81-8; ribonuclease A, 9001-99-4; lysozyme, 9001-63-2; ribonuclease T_1, 9026-12-4; staphylococcal nuclease, 9013-53-0; β-trypsin, 9002-07-7; papain, 9001-73-4; α-chymotrypsin, 9004-07-3; chymotrypsinogen, 9035-75-0; carbonic anhydrase, 9001-03-0; pepsinogen, 9001-10-9; formamide, 75-12-7; *N*-methylformamide, 123-39-7; *N*-methylacetamide, 79-16-3; *N*-ethylacetamide, 625-50-3; *N*-propylacetamide, 5331-48-6; *N*-isopropylacetamide, 1118-69-0; *N*-butylacetamide, 1119-49-9; *N*-methylpropanamide, 1187-58-2; *N*-methylbutanamide, 17794-44-4; *N*,2-dimethylpropanamide, 2675-88-9; *N*-methylpentanamide, 6225-10-1.

References

Ahmad, F., & McPhie, P. (1978) *Biochemistry 17*, 241–246.
Baker, E. N., & Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol. 44*, 97–179.
Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069–8072.
Baldwin, R. L., & Muller, N. (1992) *Proc. Natl. Acad. Sci. U.S.A.* (submitted).
Becktel, W. J., & Schellman, J. A. (1987) *Biopolymers 26*, 1859–1877.
Ben-Naim, A. (1978) *J. Phys. Chem. 82*, 792–803.
Chothia, C. (1975) *Nature 254*, 304–308.
Connelly, P., Ghosaini, L., Hu, C.-Q., Kitamura, S., Tanaka, A., & Sturtevant, J. M. (1991) *Biochemistry 30*, 1887–1891.
Dill, K. A. (1990) *Biochemistry 29*, 7133–7155.
Doig, A. J., & Williams, D. H. (1991) *J. Mol. Biol. 217*, 389–398.
Eisenberg, D., & McLachlan, A. D. (1986) *Nature 319*, 199–203.
Griko, Y. V., Privalov, P. L., Sturtevant, J. M., & Venyaminov, S. Y. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3343–3347.
Hawkes, R., Grutter, M. G., & Schellman, J. (1984) *J. Mol. Biol. 175*, 195–212.
Hawley, S. A. (1971) *Biochemistry 10*, 2436–2442.
Herzfeld, J. (1991) *Science 253*, 88.
Johnson, M. L., & Frasier, S. G. (1985) *Methods Enzymol. 117*, 301–342.
Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1–63.
Klotz, I. M., & Farnham, S. B. (1968) *Biochemistry 7*, 3879–3882.
Klotz, I. M., & Franzen, J. S. (1962) *J. Am. Chem. Soc. 84*, 3461–3466.
Konicek, J., & Wadsö, I. (1971) *Acta Chem. Scand. 25*, 1541–1551.
Kresheck, G. C., & Klotz, I. M. (1969) *Biochemistry 8*, 8–12.
Lee, B. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5154–5158.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.
Livingstone, J. R., Spolar, R. S., & Record, M. T., Jr. (1991) *Biochemistry 30*, 4237–4244.
Makhatadze, G. I., & Privalov, P. L. (1990) *J. Mol. Biol. 213*, 375–384.
Mateo, P. L., & Privalov, P. L. (1981) *FEBS Lett. 123*, 189–192.
Miller, S., Janin, J., Lesk, A. M., & Chothia, C. (1987) *J. Mol. Biol. 196*, 641–656.
Murphy, K. P., Privalov, P. L., & Gill, S. J. (1990) *Science 247*, 559–561.
Novokhatny, V. V., Kudinov, S. A., & Privalov, P. L. (1984) *J. Mol. Biol. 179*, 215–232.
Ooi, T., & Oobatake, M. (1988) *J. Biochem. (Tokyo) 104*, 440–444.
Ooi, T., & Oobatake, M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2859–2863.
Pabo, C. O., Sauer, R. T., Sturtevant, J. M., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1608–1612.
Pace, C. N. (1975) *CRC Crit. Rev. Biochem. 3*, 1–43.
Pace, C. N., & Laurents, D. V. (1989) *Biochemistry 28*, 2520–2525.
Pfeil, W. (1986) in *Thermodynamic Data for Biochemistry and Biotechnology* (Hinz, H.-J., Ed.) pp 349–376, Springer Verlag, New York.
Pfeil, W., & Privalov, P. L. (1976) *Biophys. Chem. 4*, 41–50.
Pfeil, W., & Bendzko, P. (1980) *Biochim. Biophys. Acta 626*, 73–78.
Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986) *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge.
Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167–241.
Privalov, P. L., & Khechinashvili, N. N. (1974) *J. Mol. Biol. 86*, 665–684.
Privalov, P. L., & Gill, S. J. (1988) *Adv. Protein Chem. 39*, 191–234.
Privalov, P. L., & Makhatadze, G. I. (1990) *J. Mol. Biol. 213*, 385–391.
Privalov, P. L., Tiktopulo, E. I., & Khechinashvili, N. N. (1973) *Int. J. Pept. Protein Res. 5*, 229–237.
Privalov, P. L., Mateo, P. L., Khechinashvili, N. N., Stepanov, V. M., & Revina, L. P. (1981) *J. Mol. Biol. 152*, 445–464.
Rashin, A. A. (1984) *Biopolymers 23*, 1605–1620.

Richards, F. M. (1974) *J. Mol. Biol. 82*, 1–14.
Roseman, M. A. (1988) *J. Mol. Biol. 201*, 621–623.
Salahuddin, A., & Tanford, C. (1970) *Biochemistry 9*, 1342–1347.
Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M., & Bolen, D. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2854–2858.
Sharp, K. A., Nicholls, A., Fine, R. F., & Honig, B. (1991a) *Science 252*, 106–109.
Sharp, K. A., Nicholls, A., Friedman, R., & Honig, B. (1991b) *Biochemistry 30*, 9686–9697.
Shortle, D., Meeker, A. K., & Freire, E. (1988) *Biochemistry 27*, 4761–4768.
Shrake, A., & Rupley, J. A. (1973) *J. Mol. Biol. 79*, 351–371.
Sköld, R., Suurkuusk, J., & Wadsö, I. (1976) *J. Chem. Thermodyn. 8*, 1075–1080.
Spolar, R. S., Ha, J. H., & Record, M. T., Jr. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 8382–8385.
Snedecor, G. W., & Cochran, W. G. (1980) *Statistical Methods*, Iowa State University Press, Ames, IA.
Tanford, C. (1980) *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, John Wiley and Sons, New York.
Tiktopulo, E. I., Privalov, P. L., Odintsova, T. I., Ermokhina, T. M., Krasheninnikov, I. A., Avilês, F. X., Cary, P. D., & Crane-Robinson, C. (1982) *Eur. J. Biochem. 122*, 327–331.

Promotion of the in Vitro Renaturation of Dodecameric Glutamine Synthetase from *Escherichia coli* in the Presence of GroEL (Chaperonin-60) and ATP†

Mark T. Fisher*
Laboratory of Biochemistry, National Institutes of Health, National Heart Lung and Blood Institute, Bethesda, Maryland 20892

Received November 14, 1991; Revised Manuscript Received January 24, 1992

ABSTRACT: The folding and assembly of dodecameric glutamine synthetase (GS) from *Escherichia coli* was examined in the absence and presence of the *E. coli* heat shock protein, GroEL (chaperonin-60). At nonphysiological temperatures (15–20 °C), unfolded GS spontaneously renatured to 80–90% of its original activity in the absence of GroEL. At near-physiological temperatures (37 °C), only 20–40% of the original activity returns. Under the latter solution conditions, GroEL and ATP enhance the extent of GS renaturation to 70–80% of the original activity at 37 °C. In the absence of ATP, GroEL arrests the renaturation of unfolded GS by forming a stable binary complex. The addition of ATP to this complex resulted in the release of GS subunits and formation of active dodecameric GS. The order of addition of ATP or unfolded GS to GroEL results in differences in the $t_{1/2}$ values where half-maximal GS activity is attained. At a constant GS concentration, the formation of the GroEL·GS complex followed by ATP addition resulted in approximately a 2-fold increase in the observed $t_{1/2}$ value compared to that observed when GroEL was preincubated with ATP before the GS renaturation reaction was initiated. These differences in renaturation rates may be related to binding affinity differences between the ATP-free and -bound GroEL conformer for unfolded or partially folded protein substrates [Badcoe, I. G., Smith, C. J., Wood, S., Halsall, D. J., Holbrook, J. J., Lund, P., & Clarke, A. R. (1991) *Biochemistry 30*, 9195–9200]. Although the smaller chaperonin protein, GroES (chaperonin-10), was not required for the in vitro renaturation of GS, the renaturation rates were accelerated when it was included in the reaction mixture. The addition of ATP analogues, adenosine 5′-*O*-thiomonophosphate (ATP-γ-S) and 5′-adenylylimidodiphosphate (AMP-PNP), to the arrested complex also results in the release of GS from GroEL but with slower renaturation rates and/or lower extents of GS renaturation. This suggests that ATP hydrolysis by GroEL is not required to initiate GS renaturation.

The elucidation of the physical processes which dictate protein folding has been the focus of intense research for the past three decades. Results from numerous in vitro protein folding experiments have demonstrated that extended or unfolded polypeptide chains spontaneously fold to a more compact structure. All the necessary information that is required to attain the correct folded protein structure is ultimately contained in the primary amino acid sequence. Elegant studies by King and colleagues have revealed that the primary sequence also dictates both the kinetics of formation and the folded structure of partially folded intermediates (Mitraki et al., 1991). The proper formation of these intermediates appears to be a crucial process in determining the correct folding pathway. In vivo, proteins fold in a highly complex heterogeneous environment which can potentially provide numerous interactive surfaces that can severely inhibit productive folding. The formation of folding intermediates which possess a strong tendency to form inactive aggregates usually results in the formation of inclusion bodies inside the cell. The accumulation of insoluble and/or misfolded proteins is often encountered during heat shock and overproduction of recombinant proteins (Nguyen et al., 1989; Hart et al., 1990). Although a number of oligomeric and monomeric proteins have been successfully refolded following denaturation, these reactions are typically initiated under nonphysiological solution conditions characterized by low temperatures, low protein concentrations, the

† Presented in part at the FASEB meeting, Atlanta, GA, April 21–25, 1991, and at the Fifth Symposium of the Protein Society, Baltimore, MD, June 22–26, 1991.

* Address correspondence to the author at NHLBI/NIH, Building 3, Room 207, Bethesda, MD 20892.