Thermodynamics of Interactions between Amino Acid Side Chains: Experimental Differentiation of Aromatic-Aromatic, Aromatic-Aliphatic, and Aliphatic-Aliphatic Side-Chain Interactions in Water

Antonio F. Pereira de Araujo,* Thomas C. Pochapsky,[#] and Brian Joughin[§]

*Departamento de Biologia Celular, Universidade de Brasilia, Brasilia-DF, Brazil; [#]Department of Chemistry, Brandeis University, Waltham, Massachusetts 02254 USA; and [§]Johns Hopkins University, Baltimore, Maryland, USA

ABSTRACT   A stationary phase for high-pressure liquid chromatography has been prepared by derivatizing microparticulate silica gel with functionality mimicking the side chain of isoleucine. The chromatographic retentions of a series of hydrophobic and amphiphilic amino acid analytes on this stationary phase (Ile MSP) using an aqueous mobile phase were measured as a function of temperature from 273 K to 323 K. Observed temperature dependencies are consistent with a constant change in heat capacity, ΔC_P°, upon binding of the analyte to the stationary phase. The curvatures of plots of retention data versus temperature (related to the magnitude of ΔC_P°) are distinctly different for retention of aromatic and aliphatic analytes, with retention of aliphatic analytes Val, Ile, and Leu exhibiting the characteristic signature of the hydrophobic effect, i.e., a large negative ΔC_P° upon desolvation from water and a maximum of retention around room temperature. Retention of aromatic analytes (Trp, Phe, and Tyr) involves smaller heat capacity changes and pronounced negative enthalpies of interaction with the stationary phase. Estimates of ΔC_P° for the interactions of analyte side chains with the Ile side chain were obtained by fitting the temperature dependence of retention to an expression derived from thermodynamic considerations and chromatographic theory. Similar estimates were made for interactions with the Phe side chain, using previously published data for a phenylalanine mimic stationary phase (Phe MSP) (Pochapsky and Gopen, 1992. *Protein Sci.* 1:786–795). As with the Ile MSP, the retentions of aliphatic analytes show temperature dependencies markedly different from those of aromatic analytes. Data from both phases indicate that a realistic differentiation can be made between the interactions of various types of amino acid side chains tested (i.e., aliphatic/aliphatic, aliphatic/aromatic, and aromatic/aromatic) by comparison of the corresponding thermodynamic functions for pairwise interactions. The retention of leucine on the Phe MSP and that of phenylalanine on the Ile MSP showed similar ΔC_P° values, suggesting that the aromatic-aliphatic interaction is reasonably independent of the residue attached to the stationary phase. This result is consistent with a one-to-one interaction and suggests a simple way to estimate the column-dependent phase factor, making it possible to compare entropies and free energies of interaction obtained using different MSPs. The possibilities for using MSP-derived interaction potentials in folding simulations are discussed.

INTRODUCTION

Many proteins spontaneously fold in vitro into their native conformations at the appropriate temperature and solvent composition. For small proteins this process is adequately described as a cooperative "all-or-none" transition between the denatured and native macroscopic states. For larger proteins, folding is characterized by several such cooperative transitions corresponding to the folding of discrete domains (Privalov and Gill, 1988; Griko et al., 1994). Extensive calorimetric studies in the past several decades have shown that the temperature-induced unfolding of co-operatively folding domains is accompanied by a sharp peak in heat absorption and an increase in heat capacity (a positive ΔC_P°). The free energy of unfolding has the antiintuitive property of becoming negative not only at high temperatures but also at sufficiently low temperatures. This predicted cold denaturation of proteins has been confirmed

experimentally and is now a well-established general phenomenon (Privalov et al., 1986).

Considerable progress has also been made in simulation and building a theoretical formalism for protein folding. Attempts to describe folding in terms of a succession of definite conformations of a single protein molecule, defining a folding pathway in conformational space, have given way to a statistical mechanical perspective, in which only the behavior of the ensemble is relevant (Dill and Chan, 1997; Bryngelson et al., 1995; Shakhnovich, 1997). Thermodynamic and kinetic properties of protein molecules in solution, as for any statistical system, are determined by their statistical energy landscapes. The fact that small globular proteins fold spontaneously in vitro into their native conformations under appropriate conditions of temperature and solvent implies that their energy landscapes are funnel-shaped, with folding being guided from any arbitrary conformation toward the native structure (Leopold et al., 1992; Lazaridis and Karplus, 1997).

Despite progress in developing a coherent picture of the folding process, an accurate structure prediction algorithm remains elusive (Finkelstein, 1997). It has become clear that one of the major obstacles to achieving this goal is the lack of suitable potentials for calculating the free energies of

different conformations of a given protein molecule. Conformational free energy has been expressed in some studies as a sum of pairwise interactions between amino acid residues. Numerical values for these interaction energies have been estimated from statistical analyses of available structures (Miyazawa and Jernigan, 1985; Wilson and Doniach, 1989; Kolinski et al., 1993). The basic assumptions and reliability of this procedure have been extensively debated (Thomas and Dill, 1996; Bahar and Jernigan, 1997). An alternative approach is to assume that the free energy of a given conformation can be computed from the solvent-accessible surface areas of aliphatic, aromatic, and polar residues combined with estimates of unit free energies of hydration obtained from partition experiments performed with small molecules (Makhatadze and Privalov, 1994a,b; Privalov and Makhatadze, 1993; Murphy et al., 1992; Chan and Dill, 1997).

However, no potential proposed to date has proved sufficiently accurate for realistic folding simulations or accurate ab initio structure predictions. Such a potential must be able to accurately reproduce the essential features of the energy landscape of the polypeptide. Potential inaccuracies deform the energy landscape, and if they are severe enough, will destroy the folding funnel, effectively transforming the protein into a random heteropolymer (Pereira de Araujo and Pochapsky, 1996). According to recent estimates based on the law of corresponding states between simple models and real proteins (Pereira de Araujo and Pochapsky, 1997; Onuchic et al., 1995), the accuracy of available energy functions is not far from the required limit for structure recognition (threading) experiments, but must be significantly improved for successful ab initio folding simulations.

In a previous paper (Pochapsky and Gopen, 1992), we described a novel method for quantitating the energetics of pairwise interactions between amino acid side chains, using high-pressure liquid chromatography (HPLC). Organic functionality identical to the side chain of phenylalanine was covalently bound to microparticulate silica gel, generating a HPLC stationary phase coated with a monomeric distribution of that functionality. We called this a phenylalanine side-chain mimic stationary phase **1** (Phe MSP) (see Scheme). Using water as a mobile phase, we measured the chromatographic retention of a series of N-acetyl amino acid N′-methylamides of type **2** on the Phe MSP as a function of temperature and used the resulting data to estimate the free energies of interaction between the bound and mobile side chains relative to that of glycine.

One of the important results of our earlier study was the observation that there are clear differences in the nature of the interactions of aromatic side chains and those of aliphatic side chains with the stationary phase-bound benzyl group (phenylalanine side chain) in aqueous media. On the Phe MSP, derivatives of the aromatic amino acids tryptophan and phenylalanine were far more strongly retained than those of any other amino acid at all temperatures. The logs of corrected retention factors of the aromatic amino acids (Trp, Phe, and Tyr) are essentially linear with respect
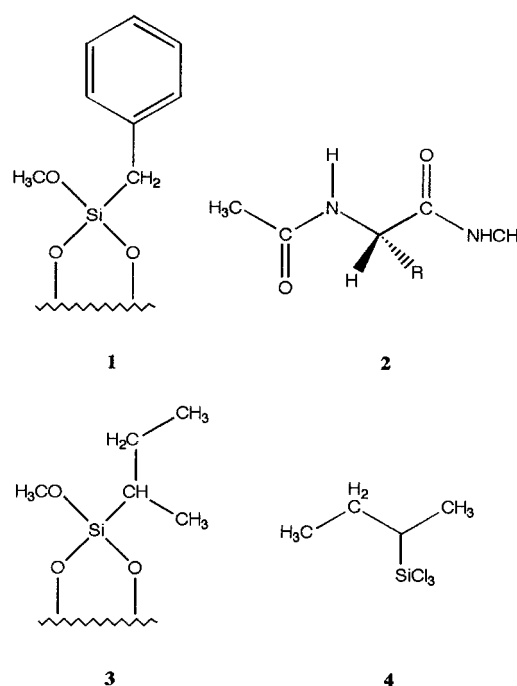
Scheme

to temperature on the Phe MSP (Fig. 1). Derivatives of the aliphatic amino acids Val and Leu are less strongly retained on the Phe MSP than the Phe and Trp derivatives at all temperatures and show larger curvatures for their temperature dependencies than the aromatic amino acids (Pochapsky and Gopen, 1992). Discrepancies between the energetics of transfer of aromatic and aliphatic compounds into water had been reported before this time, but a fundamental differentiation between the molecular mechanisms of hydration of these species that rationalize our 1992 results was only proposed more recently (Makhatadze and Privalov, 1994a,b).

Based on the considerable differences between the retention of aliphatic and aromatic amino acid derivatives on the Phe MSP, we proposed that aromatic-aromatic interactions may act as the nucleating events for general hydrophobic collapse during protein folding and would also be important in protein-protein interactions. We note that experimental work by others in recent years supports this proposal (Dadlez, 1997; Neira and Fersht, 1996; Chang et al., 1997; Lumb and Kim, 1994). We now report on the temperature dependence of the retention of type **2** analytes on an isoleucine mimic stationary phase (Ile MSP **3**) prepared by treatment of microparticulate silica gel with 2-butyltrichlorosilane **4**.

EXPERIMENTAL

Preparation of 2-(trichlorosilyl)butane 4

A 100-ml pressure-safe sealable reaction vessel was cooled to −70°C with a dry ice–acetone bath, and 3.1 g (0.055 mol) of cis-2-butene was condensed into the vessel from a gas cylinder. Distilled trichlorosilane (11.2 g,

FIGURE 1 Plot of ln k' versus T for type **2** derivatives of Gly, Val, Leu, Tyr, Phe, and Trp on the Phe MSP **1,** using water as the mobile phase. Curves connecting data points were generated by fitting experimental data to the form of Eq. 9, as described in the Experimental section. To emphasize the shapes of the curves, the fits are extrapolated above and below the experimental data, which were obtained at 273 K, 283 K, 298 K, 313 K, and 328 K. Experimental conditions and discussion of these data have been published previously (Pochapsky and Gopen, 1992). For clarity, error bars are not shown, but chromatographic data are reproducible to ±1%.
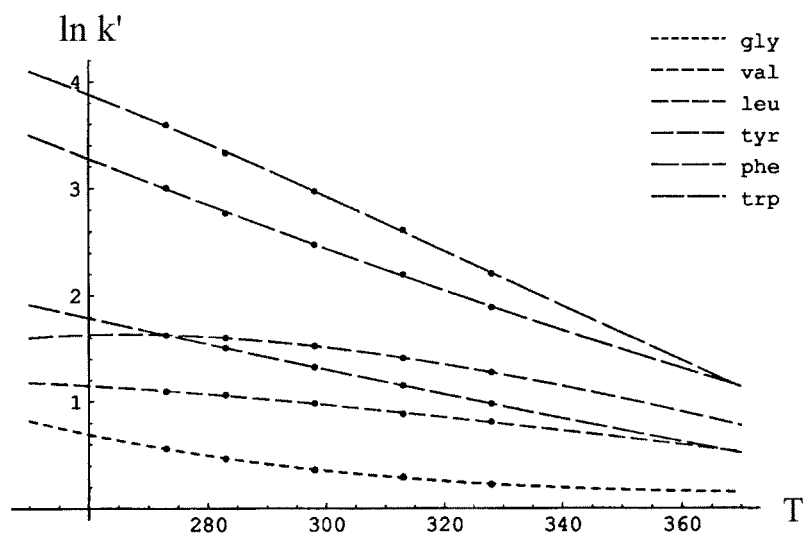
0.083 mol) was added via syringe to the vessel, as well as 0.001 g of hexachloroplatinic acid (Aldrich) dissolved in 0.05 ml of isopropyl alcohol. The vessel was sealed, allowed to warm to room temperature, and then heated slowly over several days to a final temperature of 50°C. After 5 days, the vessel was cooled and the reaction mixture was distilled under nitrogen to yield 10 g of 2-(trichlorosilyl)butane **4** (normal boiling point 57°C).

Preparation of Ile MSP 3

Five grams of 5μ silica gel (Sigma, surface area 500 m^2/g) was placed in a 250-ml three-necked round-bottom flask equipped with a mechanical stirrer. One hundred milliliters of dry toluene was added, and the silica gel was dried via azeotropic removal of water with a Dien-Stark trap until the distillate was clear. The vessel was cooled, and 1.6 g (8.4 mmol) of 2-(trichlorosilyl)butane **4** was added dropwise via syringe while the mixture was continuously stirred. Then 3.45 ml (25 mmol) of triethylamine, previously dried by distillation from CaH$_2$, was added dropwise with stirring. The reaction was heated to a gentle reflux and maintained there for 48 h. After cooling, the silica gel was separated from solvent by suction filtration on a medium frit glass filter, washed twice with methylene chloride and twice with methanol, and dried by continued suction. Loading of the bonded phase was determined by elemental analysis to be 2.8%, which corresponds to a molar loading of 0.47 mmol bonded phase per gram of silica (~180 Å^2 per molecule of bonded phase). The derivatized silica gel was packed into a standard ¼″ × 12″ stainless steel HPLC column, using a constant pressure pneumatic slurry packing pump.

Chromatographic experiments

All chromatography was performed on a Beckman System Gold HPLC system equipped with a 10-μl injection loop. For all reported experiments, deionized and vacuum degassed water was used as the mobile phase. Sample elution was detected spectrophotometrically, using the amide absorption edge at 204 nm, and recorded on a standard strip chart recorder. Type **2** analytes (N-acetyl N'-methylamides of α-amino acids) were prepared as described previously (Pochapsky and Gopen, 1992). Temperature control was obtained by immersion of the HPLC column in a stirred thermostatted water bath. Mobile phase temperature was equilibrated to the operating temperature of the column by passage through a coil of tubing also immersed in the water bath. The mobile phase flow rate was maintained at 1 ml/min for the reported data. Reported retentions (k') are corrected for column volume and were measured at the intersections of the tangents to the maximum rise and fall of the chromatographic peak.

Reported values are typically reproducible within 1% of the reported value upon repeated injection.

Data analysis

The quantity experimentally measured in a chromatographic experiment is the retention factor k', the ratio of the equilibrium concentrations of the analyte distributed between the stationary and mobile phases. Because k' changes with temperature, it is convenient to refer to this value as $k'(T)$. $k'(T)$ is obtained as the ratio of the retention time t_a of the analyte of interest corrected for the elution time of an unretained solute t_o to that of the unretained solute t_o (assuming a constant flow rate):

$$k'(T) = \frac{t_a - t_o}{t_o} \tag{1}$$

It is convenient to express $k'(T)$ as a product of a volume-independent constant $K(T)$ and the phase ratio between the volumes of the mobile phase and stationary phase ϕ:

$$k'(T) = K(T)\phi \tag{2}$$

The quantity one wishes to obtain from the measurement is the volume-independent contribution to the free energy of interaction between the solute in the mobile phase and the bound stationary phase, which is given by

$$\Delta G°(T) = -RT \ln K(T) = \Delta H° - T\Delta S° \tag{3}$$

so that

$$\ln k'(T) = \ln K(T) + \ln \phi = -\frac{\Delta H°}{RT} + \frac{\Delta S°}{R} + \ln \phi \tag{4}$$

(For a complete discussion of Eqs. 3 and 4, see Dorsey and Dill (1989).)

If both $\Delta H°$ and $\Delta S°$ are constant in the temperature range of interest, the van't Hoff plot (ln $k'(T)$ versus $1/T$) will produce a straight line with slope $\Delta H°/R$ and intercept $\Delta S°/R + \ln \phi$. If the van't Hoff plot is curved (as in most cases in the present work), it means that both the enthalpy and entropy changes due to the interaction must be treated as temperature dependent, that is, $\Delta H° = \Delta H°(T)$ and $\Delta S° = \Delta S°(T)$. An equivalent statement would be that the interaction between the bound phase and the analyte results in a nonzero change in heat capacity, $\Delta C_P°$. In the simplest case, $\Delta C_P°$ may be treated as independent of temperature over a short

temperature range. The relationships between the change in heat capacity and the thermodynamic functions are

$$\frac{\partial \Delta H^{\circ}(T)}{\partial T} = \Delta C_{P}^{\circ} \tag{5}$$

$$\frac{\partial \Delta S^{\circ}(T)}{\partial T} = \frac{\Delta C_{P}^{\circ}}{T} \tag{6}$$

Upon integration of these expressions with respect to T, we obtain

$$\Delta H^{\circ}(T) = (T - T_{max})\Delta C_{P}^{\circ} \tag{7}$$

$$\Delta S^{\circ}(T) = \Delta S^{\circ}(T_{max}) + \Delta C_{P}^{\circ} \ln \frac{T}{T_{max}} \tag{8}$$

where T_{max} is the temperature at which $\ln k'(T)$ is maximum on the van't Hoff plot and is chosen to appear in the constants of integration because $\Delta H^{\circ}(T_{max}) = 0$ (Dill, 1990). Inserting Eqs. 5 and 6 into Eqs. 7 and 8 yields

$$\ln k'(T) = \frac{T_{max}\Delta C_{P}^{\circ}}{RT} + \frac{\Delta C_{P}^{\circ}}{R} \ln T - \frac{\Delta C_{P}^{\circ}(1 + \ln T_{max})}{R}$$
$$+ \frac{\Delta S^{\circ}(T_{max})}{R} + \ln \phi \tag{9}$$

By fitting the experimental values of $\ln k'(T)$ for the retention of a particular amino acid derivative upon a given MSP to the form of Eq. 9, $A/T + B\ln T + C$, the following quantities may be obtained for each pairwise interaction:

$$B = \frac{\Delta C_{P}^{\circ}}{R} \tag{10}$$

$$\frac{A}{B} = T_{max} \tag{11}$$

$$C + B\left(1 + \ln \frac{A}{B}\right) = \frac{\Delta S^{\circ}(T_{max})}{R} + \ln \phi \tag{12}$$

The estimated value of ΔC_{P}° from Eq. 10 combined with T_{max} obtained from Eq. 11 may then be used directly in Eq. 7 to calculate the temperature dependence of the enthalpy change, $\Delta H^{\circ}(T)$. The fact that this enthalpy change can be visualized from the inclination of the van't Hoff plot follows from the derivative of Eq. 9 with respect to $1/T$:

$$\frac{\partial(\ln k')}{\partial(1/T)} = \frac{T_{max}\Delta C_{P}^{\circ}}{R} - \frac{T\Delta C_{P}^{\circ}}{R} = \frac{\Delta H_{o}(T)}{R} \tag{13}$$

Note that it is not possible to obtain an estimate of $\Delta S^{\circ}(T)$ from Eq. 8 because $\Delta S^{\circ}(T_{max})$ cannot be obtained from the van't Hoff plot. However, Eq. 12 may be used to obtain a value for the related constant $\Delta S^{\circ}(T_{max})/R + \ln \phi$, which contains the column-dependent phase ratio ϕ as an offset.

Chromatographic retentions were fit to Eq. 9 by a least-squares analysis, using Mathematica 3.0 operating on a Macintosh PowerPC 9500 or on a Gateway 2000 PC running Linux. Plots were generated using Mathematica 3.0 or MS Excel 6.0.1.

RESULTS

Although the controversy continues over the precise nature of the interactions that result in the hydrophobic effect (Makhatadze and Privalov, 1993), the hypothesis that the burying of hydrophobic residues in the interior of a protein

is a major driving force for folding is now generally accepted (Kauzmann, 1959; Baldwin, 1986; Dill, 1990; Karplus, 1997). Phenomenologically, the hydrophobic effect is recognized by a characteristic negative change in heat capacity ΔC_{P} upon desolvation of the solute (i.e., a positive change in heat capacity upon solvation), which results in a temperature dependence for both ΔH and ΔS of solvation, and a characteristic nonlinearity of the van't Hoff plot of the partition coefficients (that is, $\ln K_{P}$ versus $1/T$). We have analyzed the temperature dependence of the retention of type **2** analytes on the Ile MSP **3** to estimate the change in heat capacity upon adsorption, ΔC_{P}°, for a given analyte. ΔC_{P}° provides a quantitative measure of the differences that we noted in aliphatic and aromatic amino acid analyte retention on both the Phe and Ile MSPs and provides a useful gauge of the extent to which the interactions of different amino acids differ with a particular MSP.

Table 1 provides a comparison of the results of fitting for data obtained from the Ile MSP, with the fits graphically displayed in Fig. 2. On the Ile MSP, the largest ΔC_{P}° values are observed for the adsorption of type **2** derivatives of Val, Leu, and Ile (aliphatic-aliphatic interactions). These derivatives also show very similar values of T_{max} (between 291 and 294 K). The ΔC_{P}° values for the aromatic residues Trp, Phe, and Tyr are also clustered together and are smaller than those of the aliphatic residues. The T_{max} values for these residues are also lower, ranging from 229 K for Trp to 246 K and 248 K for Phe and Tyr. The Met derivative shows a somewhat smaller ΔC_{P}° and lower T_{max} than the aliphatic residues, reflecting perhaps the more polar nature of the side chain, whereas Gly, Ala, and Pro show small ΔC_{P}° values and lower T_{max} values as well.

Table 2 is a reanalysis of the published data for the Phe MSP, with fits graphed in Fig. 1. It is instructive to note that ΔC_{P}° values are all lower on this MSP, as are T_{max} values. Although there are as yet no direct inverse data (temperature dependence for Ile was not measured on the Phe MSP in the initial studies), the closest comparison that can be made (Leu on the Phe MSP and Phe on the Ile MSP) gives similar

TABLE 1 Estimates of T_{max}, ΔC_{P}°, and $\Delta S^{\circ}(T_{max})$ from fits of temperature dependence of retention of type 2 analytes on the Ile MSP 3, using Eqs. 10, 11, and 12 as described in the text

Type 2 analyte	T_{max} (K)	ΔC_{P}° (cal K^{-1} mol^{-1})	$\Delta S^{\circ}(T_{max}) + R \ln \phi$ (cal K^{-1} mol^{-1})
Gly	222	−5.9	0.3
Ala	268	−14.8	0.7
Val	294	−59.7	2.1
Met	272	−37.7	2.5
Pro	179	−16.9	3.7
Tyr	248	−32.3	3.2
Leu	291	−57.9	3.0
Ile	291	−48.4	3.1
Phe	246	−31.0	4.4
Trp	229	−41.7	6.0

Curves resulting from the fits are shown in Fig. 2.

FIGURE 2 Plot of ln k' versus T for type **2** derivatives of Gly, Ala, Val, Met, Pro, Tyr, Ile, Leu, Phe, and Trp on the Ile MSP **3,** using water as the mobile phase. Curves connecting data points were generated by fitting experimental data to the form of Eq. 9, as described in the Experimental section. To emphasize the shapes of the curves, the fits are extrapolated above and below the experimental data, which were obtained at 273 K, 283 K, 293 K, 303 K, 313 K, and 323 K. For clarity, error bars are not shown, but chromatographic data are reproducible to ±1%.
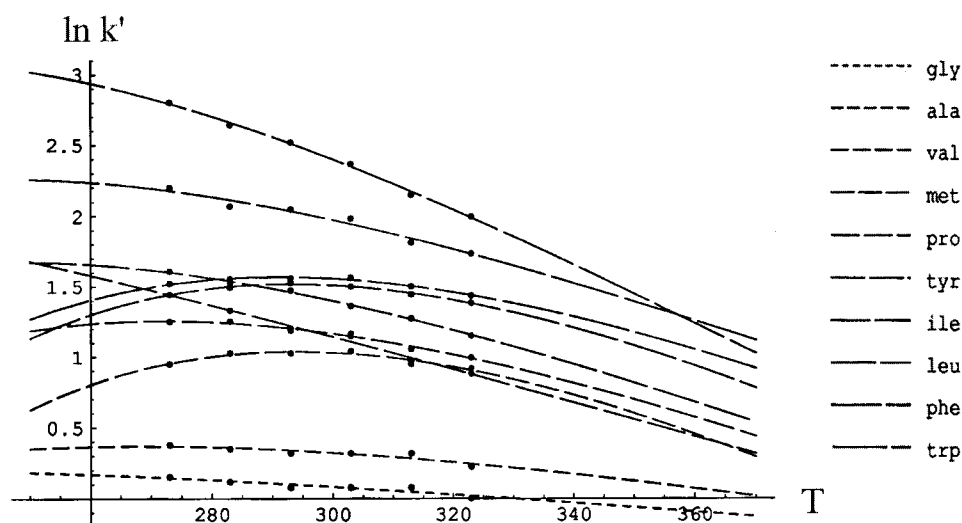
values for ΔC_P° (−34.3 cal/K mol and −31.0 cal/K mol, respectively).

The retention of aliphatic analytes on the aliphatic Ile MSP (aliphatic-aliphatic interactions) shows the characteristic signature of the hydrophobic effect, that is, a large decrease in heat capacity upon desolvation (i.e., an increase in heat capacity upon exposure of the apolar surface to water) and a maximum of retention ($\Delta H^\circ = 0$) around room temperature. Aromatic-aromatic interactions, represented by the retention of aromatic analytes on the Phe MSP (Table 1), are quite different. These involve smaller heat capacity changes and a negative interaction enthalpy throughout the experimental temperature range. Aromatic-aliphatic interactions, as represented by aromatic analyte retention on the Ile MSP and aliphatic analyte retention on the Phe MSP, lie somewhere in between.

The observation that the retention of the leucine analyte on the Phe MSP and that of the phenylalanine analyte on the Ile MSP show similar changes in heat capacity and extrapolated retention maxima suggests that the thermodynamics of these interactions are relatively independent of which side chain is attached to the stationary phase. This result is encouraging, as it is consistent with one-to-one interactions between side chains on the MSP, and it is a mandatory

condition for any use of these data as parameters in folding simulations. Although no temperature-dependent data for isoleucine retention on the Phe MSP was collected in our previous study, the fact that the isoleucine and leucine retentions on the Ile MSP are virtually identical suggests that they might also be similar on the Phe MSP. Furthermore, although Phe retention on the Ile MSP and Leu retention on the Phe MSP show similar changes in heat capacity, they differ in absolute retention. This indicates that the observed difference in retention at a given temperature is likely to be due only to differences in the column-dependent phase factor ϕ, allowing an estimate of this number to be made. Using Eqs. 10–12 and the estimated value of ϕ, it should be possible to compare not only enthalpies, but entropies and free energies of interaction obtained using different MSPs.

Enthalpies of interaction obtained using Eq. 7 are shown in Fig. 3 for the Ile MSP and in Fig. 4 for the Phe MSP. Although extrapolated enthalpy values are shown in these figures down to 250 K and up to 410 K, it is important to note that the experimental measurements from which these plots are derived were performed over a temperature range of 273 K to 323 K, and that small uncertainties in the data might be significantly amplified in the extrapolated region, even if the assumption of a constant ΔC_P° remains valid over the entire temperature range.

All measured pairwise interactions exhibit a negative $\Delta H^\circ(T)$ at 300 K, that is, they are enthalpically favored. At this temperature, $\Delta H^\circ(T)$ is most negative for aromatic-aromatic interactions, as represented by the retention of the Trp and Phe derivatives on the Phe MSP ($\Delta H_{300}^\circ = -9.0$ and $\Delta H_{300}^\circ = -7.2$ kcal/mol for Trp and Phe on the Phe MSP, respectively). It is the least negative for aliphatic-aliphatic interactions, represented by the retentions of the Val, Leu, and Ile derivatives on the Ile MSP, where $\Delta H^\circ(T)$ reaches 0 near 293 K and becomes positive (unfavorable) below this temperature.

TABLE 2 Estimates of T_{max}, ΔC_P°, and $\Delta S^\circ(T_{max})$ from fits of temperature dependence of retention of type 2 analytes on the Phe MSP 1, using Eqs. 10, 11, and 12 as described in the text

Type **2** analyte	T_{max} (K)	ΔC_P° (cal K^{-1} mol^{-1})	$\Delta S^\circ(T_{max}) + R \ln \phi$ (cal K^{-1} mol^{-1})
Gly	380	13.1	0.3
Val	224	−12.8	2.4
Leu	265	−34.3	3.2
Tyr	101	−10.6	7.1
Phe	75	−16.0	14.9
Trp	176	−36.1	10.1

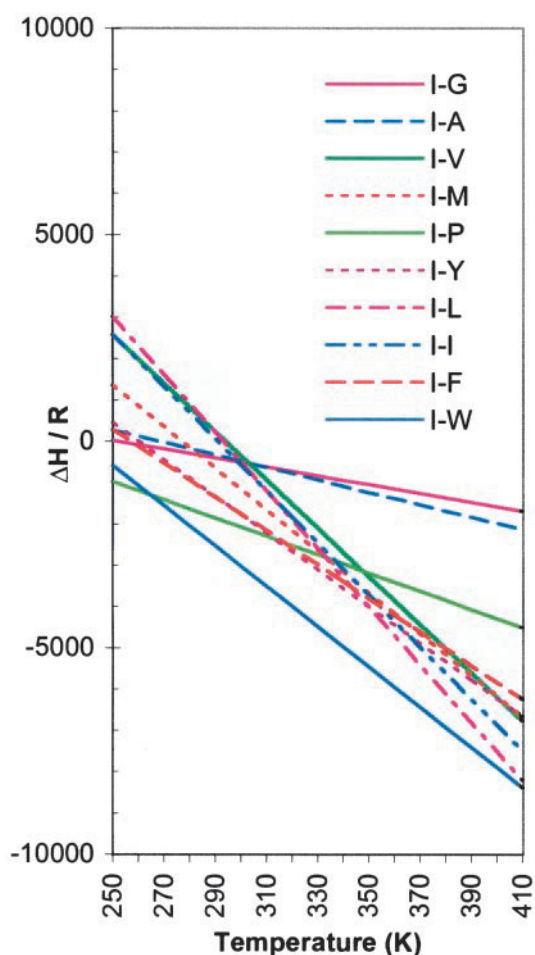Curves resulting from the fits are shown in Fig. 1.

FIGURE 3 Plot of $\Delta H°(T)/R$ versus T for type **2** derivatives of Gly, Ala, Val, Met, Pro, Tyr, Ile, Leu, Phe, and Trp on the Ile MSP **3**, using water as the mobile phase. Curves were generated with Eq. 8, using the values of $\Delta C_P^°$ and T_{max} shown in Table 2. To emphasize the shapes of the curves, they are extrapolated above and below the experimental range from 273 K to 323 K.

FIGURE 4 Plot of $\Delta H°(T)/R$ versus T for type **2** derivatives of Gly, Val, Leu, Tyr, Phe, and Trp on the Phe MSP **1**, using water as the mobile phase. Curves were generated with Eq. 8, using the values of $\Delta C_P^°$ and T_{max} shown in Table 1. To emphasize the shapes of the curves, they are extrapolated above and below the experimental range (273 K to 328 K) from 250 K to 410 K.

Scaled entropies of interaction on both MSPs are shown in Figs. 5 and 6. Entropy values obtained using Eqs. 8 and 12 necessarily include the column-dependent phase factor ln ϕ, as discussed above. However, the difference in the entropies for different pairwise interactions on a given MSP correspond to the difference between the entropic contributions of the respective pairwise interactions, because the phase factor is a constant for a given column. On the Phe MSP at 300 K, the most negative interaction entropy is observed for the Trp derivative, followed by Phe and Tyr. The most positive interaction entropies correspond to the Leu and Val derivatives, followed by the Gly derivative. Clearly, the interactions of aromatic residues with the Phe MSP are entropically less favorable than the interactions of aliphatic residues, and the longer retentions of the aromatic residues are due to favorable enthalpies.

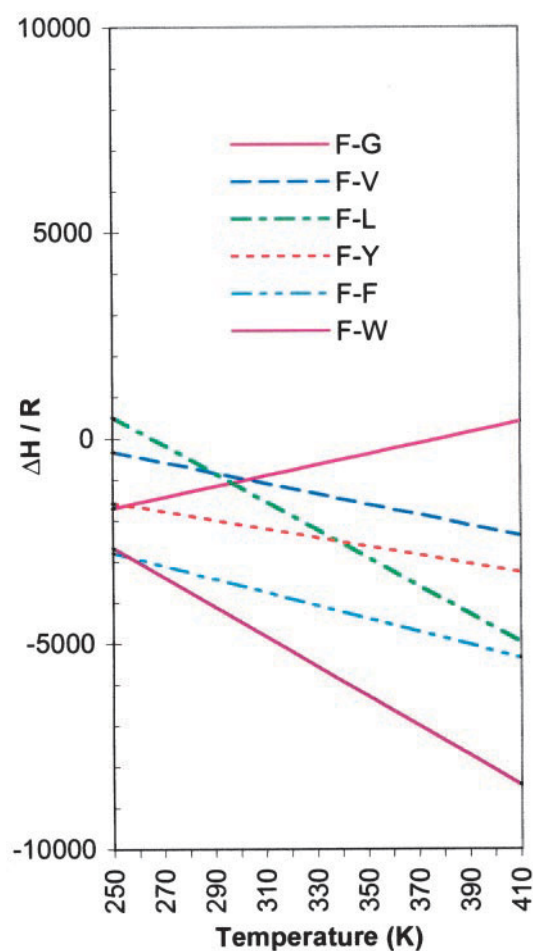The results described here for both MSPs suggest that aliphatic-aliphatic interactions in water are typically hydrophobic, with large negative changes in heat capacities and near-zero interaction enthalpies near room temperature. The favorable entropy that drives the interaction likely results from disordering of solvent. Any decrease in side-chain conformational entropy that results from the interaction is small and is surpassed by the increase in solvent entropy. The two interacting side chains may well maintain a high conformational entropy, interconverting between many configurations as if they were in a liquid phase. The smaller heat capacity changes, very favorable enthalpies, and less favorable entropies indicate a different mechanism for aromatic-aromatic interactions. This is somewhat unexpected, because the temperature dependence of solution of benzene in water shows a typical hydrophobic profile (Tanford, 1980), and one might expect that the retention of the Phe derivative on the Phe MSP would result in a similar profile. One possible explanation for this apparent contradiction is that interacting aromatic functionality in the Phe-Phe pairing does not behave as benzene molecules do in a liquid, where interactions are not one to one and there is little or no
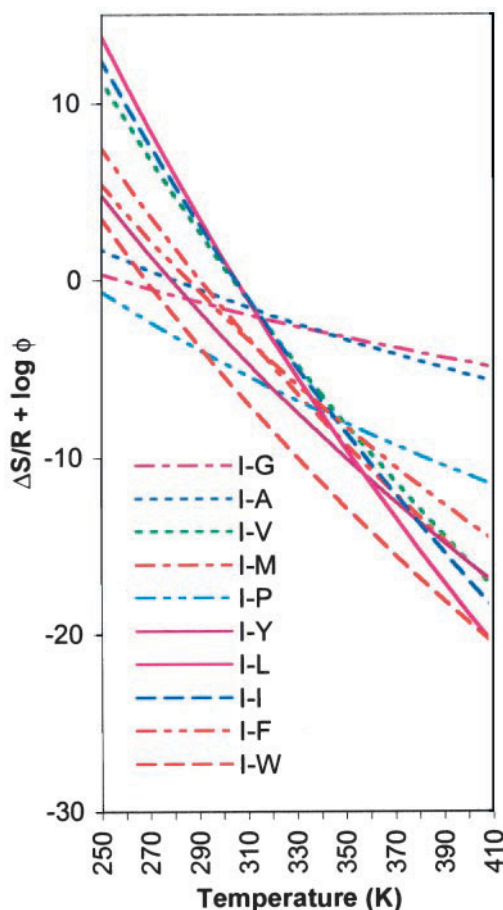
FIGURE 5 Plot of $\Delta S°(T)/R + \ln \phi$ versus T for type **2** derivatives of Gly, Ala, Val, Met, Pro, Tyr, Ile, Leu, Phe, and Trp on the Ile MSP **3,** using water as the mobile phase. Curves were generated with Eq. 9, using the values of $\Delta C_P°$ and T_{max} shown in Table 2. To emphasize the shapes of the curves, they are extrapolated above and below the experimental range (273 K to 328 K) from 250 K to 410 K.

FIGURE 6 Plot of $\Delta S°(T)/R + \ln \phi$ versus T for type **2** derivatives of Gly, Val, Leu, Tyr, Phe, and Trp on the Phe MSP **1,** using water as the mobile phase. Curves were generated with Eq. 9, using the values of $\Delta C_P°$ and T_{max} shown in Table 1. To emphasize the shapes of the curves, they are extrapolated above and below the experimental range (273 K to 328 K) from 250 K to 410 K.

steric demand on intermolecular interactions. Instead, one-to-one aromatic-aromatic interactions observed using the Phe MSP may be more like interactions found in a solid, in which a few enthalpically favorable relative configurations predominate. This also rationalizes the rather small heat capacity changes observed for aromatic-aromatic interactions, because solid-like interactions would increase the heat capacity of the interacting side chains while leaving the heat capacities of the solvated side chains unaffected.

DISCUSSION

Comparison of MSP results with small-molecule partition experiments

Our results indicate that aliphatic-aliphatic interactions in water are typically hydrophobic, with large negative changes in heat capacities and negligible interaction enthalpies between 290 K and 300 K. These results are in excellent agreement with transfer experiments of pure liquid hydrocarbons into water (Privalov and Gill, 1988). The
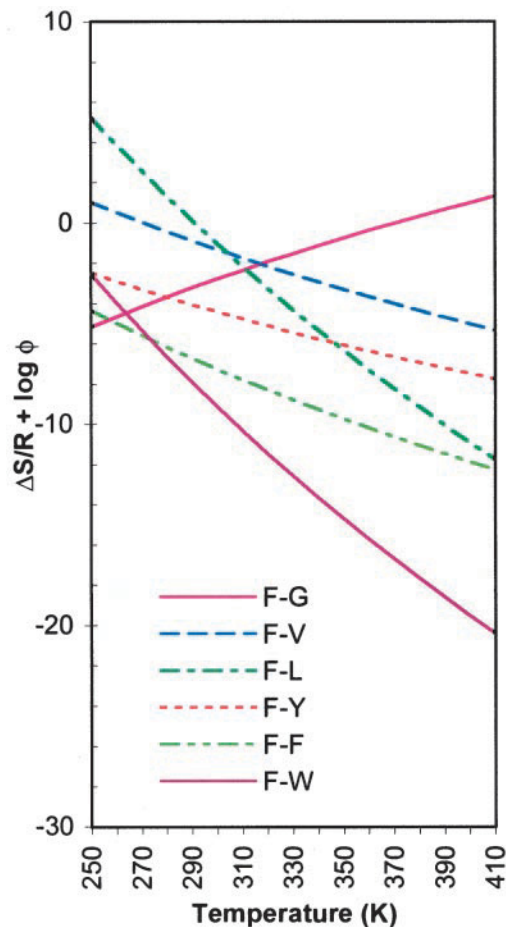
partition measurements also give an extrapolated characteristic temperature for a zero entropy change around 380 K, assuming a constant $\Delta C_P°$. Comparison with the scaled entropies for aliphatic-aliphatic interactions obtained from measurements made on the Ile MSP for Val, Leu, and Ile may provide an estimate for the log of the phase factor (ln ϕ) of the Ile MSP between -13 and -15 at 380 K.

The observed magnitude of the heat capacity change for aliphatic-aliphatic interactions on the Ile MSP also falls in a range reasonable for one-to-one interactions between the MSP-bound side chain and that of the analyte (as opposed to nonspecific desolvation of the analyte). The $\Delta C_P°$ for Ile-Ile and Ile-Leu interactions is approximately -130 cal/K mol (-550 J/K mol). The temperature dependence of the enthalpy of hydration per unit of water-accessible aliphatic surface (Makhatadze and Privalov, 1993) provides for an estimated heat capacity change between 2.2 and 2.6 J/K mol Å^2. Using a value of $\Delta C_P°$ for vaporization of the pure hydrocarbon of -0.2 J/K mol Å^2 (Makhatadze and Privalov, 1993), we arrive at an estimate between 2.0 and

2.4 J/K mol Å^2 for ΔC_P° per unit aliphatic surface transferred from the pure hydrocarbon phase to water. Based on current experimental data and these estimates, the surface involved in aliphatic-aliphatic contact on the Ile MSP for Ile-Ile and Ile-Leu is between 229 and 275 Å^2. Considering that some of the heat capacity change may be due to the peptide moiety (which may be as large as −89 J/K mol, based on the retention of the Gly analyte on the Ile MSP), the total aliphatic surface protected from water in the contact may well be smaller than the total surface area of a single Ile side chain (223 Å^2, according to Sharp et al., 1991).

On the other hand, the behavior of aromatic analytes on the Phe MSP cannot be completely understood from what is known about the partition of aromatic species between phases. It is true that the transfer of an aromatic species from a pure phase to water is accompanied by a smaller heat capacity change than the same process for aliphatics, but the change is still significant and, furthermore, exhibits an enthalpy near zero at ambient temperatures, as do aliphatic compounds (Tanford, 1980; Privalov and Gill, 1988). The smaller heat capacity change for aromatic transfer from pure phase into water at 298 K is due almost entirely to the difference between ΔC_P upon hydration of aliphatic and aromatic compounds, as the heat capacity changes upon vaporization are much smaller and quite similar for benzene and cyclohexane (Makhatadze and Privalov, 1993). These workers also indicate that the heat capacity change upon hydration is ~1.4 times as great per unit area for aliphatic compounds as for aromatic species. However, the heat capacity changes obtained from the MSP data for Ile-Leu or Ile-Ile are ~4 times larger than that for Phe-Phe. Furthermore, there is no agreement between the T_{max} values measured for the partition of simple aromatic compounds into water (near ambient temperature) versus those extrapolated from chromatographic data (~75 K for the Phe-Phe pairing).

The observed differences between aliphatic-aliphatic and aromatic-aromatic interactions may be at least partially explained by the recently proposed differences in their respective mechanisms for molecular hydration (Makhatadze and Privalov, 1994a). Based on thermodynamic measurements for transfer of aromatic and aliphatic compounds into water, these authors concluded that hydration of aromatic compounds is thermodynamically favorable, corresponding to a negative change in free energy upon solution in water, whereas the opposite is true for aliphatic hydrocarbons. For aliphatic compounds, the process is dominated by the unfavorable entropy decrease classically associated with the ordering of water molecules around the hydrophobic surface. For aromatic compounds, on the other hand, the entropy decrease upon transfer from the gas phase to water is not as pronounced and the free energy of solvation is dominated by a favorable enthalpy decrease. A less ordered solvation shell, as suggested by the smaller entropy decrease upon solvation, might also explain the smaller increase in heat capacity upon hydration of aromatic compounds relative to aliphatic species. Based on their observations, the

authors concluded that the well-known low solubility of liquid aromatics in water was due to favorable interactions between the aromatic molecules in the pure liquid phase, not to unfavorable interactions of the aromatic compounds with water (Makhatadze and Privalov, 1994a).

Hydration considerations can qualitatively rationalize the smaller change in heat capacity observed for aromatic-aromatic interactions relative to aliphatic-aliphatic ones. The reduced importance of entropy in aromatic-aromatic interactions relative to aliphatic-aliphatic interactions is also understandable if the ordering of water molecules around aromatic molecules is less than for aliphatic species. It is clear, however, that the very favorable enthalpies of interaction observed for aromatic-aromatic pairwise interactions with MSP **1** cannot be explained by dehydration, because the enthalpy of hydration of these species is negative, and so its contribution to the contact enthalpy should be positive. Based on these considerations, the pairwise interaction between aromatic side chains must be quite exothermic and could be expected to differ from the nonpairwise interactions available in pure liquid aromatic hydrocarbon in some fundamental ways.

Chromatography

In our previous paper, we discussed at length the justification for considering the retention of type **2** analytes on an MSP as a faithful model for the interactions that occur between amino acid side chains in an aqueous environment (Pochapsky and Gopen, 1992). Basic to our argument is the assumption that the interactions between bonded phase and analyte are primarily 1:1 between individual side chains, and that MSP retention does not reflect interaction between the analyte and many bonded phase side chains simultaneously (adsorption rather than partition). There are several reasons to believe that the observed values are indeed due primarily to 1:1 interactions between the bound and analyte side chains. First, surface loading calculations lead to an estimate of 80–160 Å^2 surface area per molecule of bound side chain on the MSPs prepared so far. This translates into an average spacing of 10–14 Å between bonded side chains. This calculation agrees well with data obtained from chromatography of bidentate ligands, which also indicate an average spacing of 10–15 Å on stationary phases prepared using polychlorosilanes by methods similar to those used to prepare MSP HPLC phases (Pirkle and Pochapsky, 1988). Even the largest amino acid side chains are significantly smaller than this, so although measurements made using MSPs would be unlikely to be entirely free of cooperativity effects, adsorption should consist mostly of 1:1 interactions.

Short-chain bonded phases for HPLC retain analytes primarily via adsorption, unlike the long-chain bonded phases such as C18, for which partition is the primary retention mechanism (Dill, 1987; Dorsey and Dill, 1989). Adsorptive behavior is primarily a surface phenomenon and precludes a significant degree of cooperative binding. Adsorptive bind-

ing also requires desolvation of the analyte and bound functionality only at the interface between them, so the magnitude of free energy changes upon adsorption is therefore significantly smaller than for partition. A simple lattice model predicts reduction of the free energy change by a factor of ⅙ (Dorsey and Dill, 1989). A comparison of retention of analytes of type **2** on MSPs prepared so far and retention of the same analytes on C18 HPLC phases confirms that the energetics of MSP retention for type **2** analytes are of the order of magnitude expected for adsorptive behavior (Pochapsky and Gopen, 1992). The ease with which MSP bonded phases are overloaded by large samples of analyte is consistent with an adsorptive mechanism.

Incorporation of MSP-HPLC data into residue-specific interaction potentials and comparison with other parameterization methods

The long-term goal of this research is to prepare a table of pairwise-specific potentials for interactions between amino acid side chains for use in folding simulations and other applications. The simplest approach to such a set of potentials is to measure k' relative to a standard (Gly being the obvious choice) and use the relative retentions to derive $\Delta\Delta G_{iG}$ for each side chain i according the expression $\Delta\Delta G_{iG} = -RT \ln \alpha_{iG}$, where $\alpha_{iG} = k'_i/k'_{Gly}$. This approach was used to compare pairwise interactions in our earlier paper (Pochapsky and Gopen, 1992). However, to generate a set of pairwise-specific potentials using such values obtained from different MSPs, it would be necessary to normalize the raw thermodynamic data from each MSP, making the choice of a reference state important. The difficulty is that despite the lack of a side chain, Gly may interact differently with different amino acids in ways that the other amino acids could not precisely duplicate for steric reasons. For example, one can imagine efficient simultaneous pi-stacking between Trp and both amide groups of the Gly derivative. This would invalidate or at least render quite difficult the use of a Gly standard. An alternative approach is suggested by the results presented here, that is, the similarity in the ΔC_P values and T_{max} for the Leu derivative on the Phe MSP and the Phe derivative on the Ile MSP. Assuming that inverse interaction pairings (e.g., Leu on the Phe MSP and Phe on the Leu MSP) are by definition equivalent, a value can be estimated for the phase ratio ϕ for a given MSP as described above. Once the phase ratio has been estimated, values for the $\Delta H°(T)$ and $\Delta S°(T)$ of pairwise interactions may also be calculated. This approach is particularly attractive in that it uses the solvated form of both side chains as a reference state, which is the most useful reference state for the initial stages of protein folding.

It is appropriate to consider what advantages MSP-derived parameterizations might have to offer compared to other commonly used methods for estimating the interaction energies of amino acid side chains. Such methods include potentials extrapolated from hydrophobicity measurements and potentials obtained by statistical analysis of contacts within protein structures. A number of hydrophobicity scales have been determined experimentally (Nozaki and Tanford, 1971; Damodaran and Song, 1986; Fauchere and Pliska, 1983; Radzicka and Wolfenden, 1988). The interpretation of such scales and their appropriateness for use in characterizing specific interactions involved in folding and recognition have also been the subject of considerable effort, particularly because many of the scales show conflicting results (Karplus, 1997). However, because the MSP method directly examines the interactions between amino acid side chains, the problems of extracting the desired information (energetics of side chain–side chain interactions) from measured partition coefficients, which contain translational and conformational entropy terms, are largely avoided (Chan and Dill, 1997).

Another common method for calculating interaction (contact) free energies is based on statistical analysis of pairwise side-chain contacts in folded proteins (Tanaka and Scheraga, 1976; Miyazawa and Jernigan, 1985, 1996). Such potentials are calculated by considering the statistical frequency of pairwise contact between a particular i and j residue pair in the structure database relative to some reference state as a Boltzmann distribution (the quasichemical approximation). This allows the calculation of a contact energy for the i-j pair relative to the reference state. The various scales differ in such features as reference state, how contacts are defined, how the contact potentials are corrected for the effects of chain connectivity, side-chain size, solvent contacts, and relative abundance of each residue. Because of their widespread use in threading, structure evaluation, and folding simulations, knowledge-based potentials have been subjected to considerable scrutiny to determine their strengths and weaknesses (Ben-Naim, 1997). Thomas and Dill tested the assumption of independence of specific pairwise interactions in the database, using exhaustive enumeration of conformations in simple lattice models (Thomas and Dill, 1996). They found the form of distance-dependent potentials, even for real proteins, could be reproduced with considerable accuracy by simply assuming a hydrophobic potential that places hydrophobic residues in the protein interior. Furthermore, extracted potentials were found to depend on chain length. Finally, there is some question about the appropriateness of using Boltzmann distribution laws to extract contact energies from contact frequencies in the database, because the PDB database is fixed, and so the appropriate temperature for considering these interactions is difficult to establish.

Perhaps the most important advantage that the MSP-HPLC method offers is that the interactions being measured are almost precisely those that might be expected to occur in the initial stages of protein folding: isolated or near-isolated contacts between amino acid side chains. Partition between an organic solvent and water does not accurately reflect the energetics of specific interactions between amino acid side chains, because partition free energies contain translational and conformational entropy terms (Chan and Dill, 1997).

Furthermore, the enthalpic and entropic changes that arise from the complete desolvation of an amino acid (as expected upon partition into an organic phase) should be quite different from the partial desolvation of a side chain that occurs during the interaction between two side chains during the early stages of folding.

The ability to extract ΔC_P° values for pairwise interactions is another attractive feature of the MSP approach. Using Eqs. 7 and 8, the explicit temperature dependence of the interactions can be incorporated directly into the pairwise potentials. Most current methods for parameterizing side-chain interactions do not offer this option; free energies of interaction are assumed to be temperature independent. Because the temperature dependence of the potential is critical to the behavior of real proteins (Privalov et al., 1986), this is a major advantage of the MSP-HPLC method over virtually every other method for measuring such interactions, except direct calorimetry.

REFERENCES

Bahar, I., and R. L. Jernigan. 1997. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* 266:195–214.

Baldwin, R. L. 1986. Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. USA.* 83:8069–8072.

Ben-Naim, A. 1997. Statistical potentials derived from protein structures: are these meaningful potentials? *J. Chem. Phys.* 107:3698–3706.

Bryngelson, J. D., J. N. Onuchic, N. D. Socci, and P. G. Wolnyes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins Struct. Funct. Genet.* 21:167–195.

Chan, H. S., and K. A. Dill. 1997. Solvation: how to obtain microscopic energies from partitioning and solvation experiments. *Annu. Rev. Biophys. Biomol. Struct.* 26:425–459.

Chang, Y., J. Zajicek, and F. J. Castellino. 1997. Role of tryptophan-63 of the kringle 2 domain of tissue-type plasminogen activator in its thermal stability, folding and ligand binding properties. *Biochemistry.* 36:7652–7663.

Dadlez, M. 1997. Hydrophobic interactions accelerate early stages of the folding of BPTI. *Biochemistry.* 36:2788–2797.

Damodaran, S., and K. B. Song. 1986. The role of solvent polarity in the free energy of transfer of amino acid side chains from water to organic solvents. *J. Biol. Chem.* 261:7220–7222.

Dill, K. A. 1987. The mechanism of solute retention in reversed-phase liquid chromatography. *J. Phys. Chem.* 91:1980–1988.

Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry.* 29:7133–7155.

Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nature Struct. Biol.* 4:10–19.

Dorsey, J. G., and K. A. Dill. 1989. The molecular mechanism of retention in reversed-phase liquid chromatography. *Chem. Rev.* 89:331–345.

Fauchere, J.-L., and V. Pliska. 1983. Hydrophobic parameters π of amino acid side chains from the partitioning of N-acetyl-amino acid amides. *Eur. J. Med. Chem.* 18:369–375.

Finkelstein, A. V. 1997. Protein structure: what is possible to predict now? *Curr. Opin. Struct. Biol.* 7:61–71.

Griko, Y. V., A. Gittis, E. E. Lattman, and P. L. Privalov. 1994. Residual structure in a staphylococcyl nuclease fragment: is it a molten globule and is its unfolding a first-order phase transition? *J. Mol. Biol.* 243:93–99.

Karplus, P. A. 1997. Hydrophobicity regained. *Protein Sci.* 6:1302–1307.

Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63.

Kolinski, A., A. Godzik, and J. Skolnick. 1993. A general method for the prediction of the three dimensional structure and folding pathway of globular proteins: application to designed helical proteins. *J. Chem. Phys.* 98:7420–7433.

Lazaridis, T., and M. Karplus. 1997. New view of protein folding reconciled with the old through multiple unfolding simulations. *Science.* 278:1928–1931.

Leopold, P. E., M. Montal, and J. N. Onuchic. 1992. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA.* 89:8721–8725.

Lumb, K. J., and P. S. Kim. 1994. Formation of a hydrophobic cluster in denatured bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* 236:412–420.

Makhatadze, G. I., and P. I. Privalov. 1993. Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J. Mol. Biol.* 232:639–659.

Makhatadze, G. I., and P. I. Privalov. 1994a. Energetics of interactions of aromatic hydrocarbons with water. *Biophys. Chem.* 50:285–291.

Makhatadze, G. I., and P. I. Privalov. 1994b. Hydration effects in protein folding. *Biophys. Chem.* 50:291–309.

Miyazawa, S., and R. L. Jernigan. 1985. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules.* 18:534–552.

Miyazawa, S., and R. L. Jernigan. 1996. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* 256:623–644.

Neira, J. L., and A. R. Fersht. 1996. An NMR study on the beta-hairpin region of barnase. *Fold. Des.* 1:231–241.

Nozaki, Y., and C. Tanford. 1971. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J. Biol. Chem.* 246:2211–2217.

Murphy, K. P., V. Bhakuni, D. Xie, and E. Friere. 1992. Molecular basis of cooperativity in protein folding. III. Structural identification of cooperative folding units and folding intermediates. *J. Mol. Biol.* 227:293–306.

Onuchic, J., P. Wolynes, Z. Luthey-Schulten, and N. Socci. 1995. Toward the outline of the topography of a realistic protein folding funnel. *Proc. Natl. Acad. Sci. USA.* 92:3626–3630.

Pereira de Araujo, A. F., and T. C. Pochapsky. 1996. Monte Carlo simulations of protein folding using inexact potentials: how accurate must parameters be in order to preserve the essential features of the energy landscape? *Fold. Des.* 1:299–314.

Pereira de Araujo, A. F., and T. C. Pochapsky. 1997. Estimates for the potential accuracy required in realistic protein folding simulations and structure recognition experiments. *Fold. Des.* 2:135–139.

Pirkle, W. H., and T. C. Pochapsky. 1988. Separation of the stereoisomers of a homologous series of bis-amides on chiral stationary phases. *Chromatographia.* 25:652–654.

Pochapsky, T. C., and Q. Gopen. 1992. A chromatographic approach to the determination of relative free energies of interaction between hydrophobic and amphiphilic amino acid side chains. *Protein Sci.* 1:786–795.

Privalov, P. L., and S. J. Gill. 1988. Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* 39:191–234.

Privalov, P. L., Yu. V. Griko, and S. Yu. Venyaminov. 1986. Cold denaturation of myoglobin. *J. Mol. Biol.* 190:487–498.

Radzicka, A., and R. Wolfenden. 1988. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry.* 27:1664–1670.

Shakhnovich, E. I. 1997. Theoretical studies of protein folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* 7:29–40.

Sharp, K. A., A. Nicholls, R. Friedman, and B. Honig. 1991. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry.* 30:9686–9687.

Tanaka, S., and H. A. Scheraga. 1976. Medium and long-range interaction parameters between amino acids for predicting three dimensional structures of proteins. *Macromolecules.* 9:945–950.

Tanford, C. 1980. The Hydrophobic Effect, 2nd ed. Wiley, New York.

Thomas, P. D., and K. A. Dill. 1996. Statistical potentials extracted from protein structures: how accurate are they? *J. Mol. Biol.* 257:457–469.

Wilson, C., and S. Doniach. 1989. A computer model to dynamically simulate protein folding: studies with crambin. *Proteins.* 6:193–209.