Energetic Origins of Specificity of Ligand Binding in an Interior Nonpolar Cavity of T4 Lysozyme†

Andrew Morton, Walter A. Baase, and Brian W. Matthews*

Institute of Molecular Biology, Howard Hughes Medical Institute, and Departments of Chemistry and Physics, University of Oregon, Eugene, Oregon 97403

Received February 9, 1995; Revised Manuscript Received April 21, 1995®

ABSTRACT: To determine the constraints on interactions within the core of a folded protein, we have analyzed the binding of 91 different compounds to an internal cavity created in the interior of phage T4 lysozyme by site-directed mutagenesis [Eriksson et al. (1992a) *Nature 355*, 371−373]. The cavity is able to accommodate a variety of small, mainly nonpolar, ligands. Molecules which do not appear to bind include those that are very polar, those that are too large, and those that have appropriate volume and polarity but inappropriate shape. Calorimetric analysis of 16 of these ligands reveals that their free energies of binding are poorly correlated with their solvent-transfer free energies. In addition, their enthalpies of binding are much larger than expected on the basis of transfer of the ligands from an aqueous to a nonpolar liquid phase. The binding energetics were analyzed by dividing the reaction into three processes: desolvation, immobilization, and packing. This analysis indicates that all three processes contribute to binding specificity. For a subset of these ligands that are structurally related, however, packing interactions in the protein interior are well modeled by the interactions of the ligands with octanol.

The multiple configurations available to a protein and the complexity of interactions both within the protein and between the protein and solvent have precluded direct calculation of the relative stability of the protein configuration. An important simplification has been the use of small molecule solvent-transfer free energies to model the free energy associated with removing nonpolar groups from water during folding (Kauzmann, 1959; Dill, 1990). While such models capture the effect of nonpolar groups on the structure of aqueous solutions, solvent-transfer measurements alone are either inappropriate or inadequate models of other important aspects of the energetics of nonpolar associations. Protein interiors are densely packed, for example, and are more rigid than organic liquids (Richards, 1974). Results from directed mutagenesis experiments [e.g., Yutani et al. (1987), Matsumura et al. (1988), Kellis et al. (1989), Karpusas et al. (1989), Lim and Sauer (1989), Shortle et al. (1990), Sandberg and Terwilliger (1991), Eriksson et al. (1992b), and Baldwin et al. (1993)] demonstrate that specific geometric or packing interactions which vary from site to site and from protein to protein can provide a significant contribution to protein stability. Because of this context dependence, packing interactions must be considered separately from hydrophobic effects. Also the reduction in conformational entropy associated with protein folding does not occur upon solvent transfer of model compounds. Entropic effects must therefore be considered separately from the hydrophobic effect. This is also the case for protein−ligand associations.

The central problem, then, in using the concept of hydrophobicity to understand protein stability is to account for the effects of additional factors such as site-specific packing interactions and configurational entropy which are not well modeled by the transfer of a nonpolar molecule from water to organic solvent. To address this problem, one must first identify such factors and then find an experimental way to decouple them from the hydrophobic effect.

Conventional mutational analyses cannot readily distinguish hydrophobic effects from packing effects. Given the 20 naturally occurring amino acids, it is difficult to incrementally alter hydrophobicity without altering shape and *vice versa*. For this reason, it is difficult to assign observed changes in structure or energy upon mutation to either steric or hydrophobic origins. Ligand binding studies are more versatile than mutational studies in this respect, since different ligands can be used which have very similar shapes but different hydrophobicities (Connelly et al., 1990) or similar hydrophobicities but different shapes. Most enzyme−ligand complexes, however, involve specific polar and electrostatic interactions that are easily perturbed by subtle alterations in ligand shape or polarity. Their energetics are further complicated in that the ligand and the binding site typically become only partially desolvated upon association. The simplest system with which to assess the relation between hydrophobic and steric energies would thus involve a totally buried, nonpolar binding site.

A system which meets these criteria has been described in T4 lysozyme, where a leucine to alanine mutation at position 99 (L99A) in the protein core results in expansion of a small preexisting cavity into one capable of binding a benzene molecule (Eriksson et al., 1992a). This report summarizes the association of 91 different compounds with this cavity-containing mutant. Quantitative binding data have been obtained for a selected subset of these compounds. Crystal structures of the protein−ligand complexes are described in the accompanying paper (Morton & Matthews, 1995).

† A.M. was a Howard Hughes Predoctoral Fellow. This work was supported in part by NIH Grant GM21967 to B.W.M.

* Author to whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* June 15, 1995.

0006-2960/95/0434-8564$09.00/0

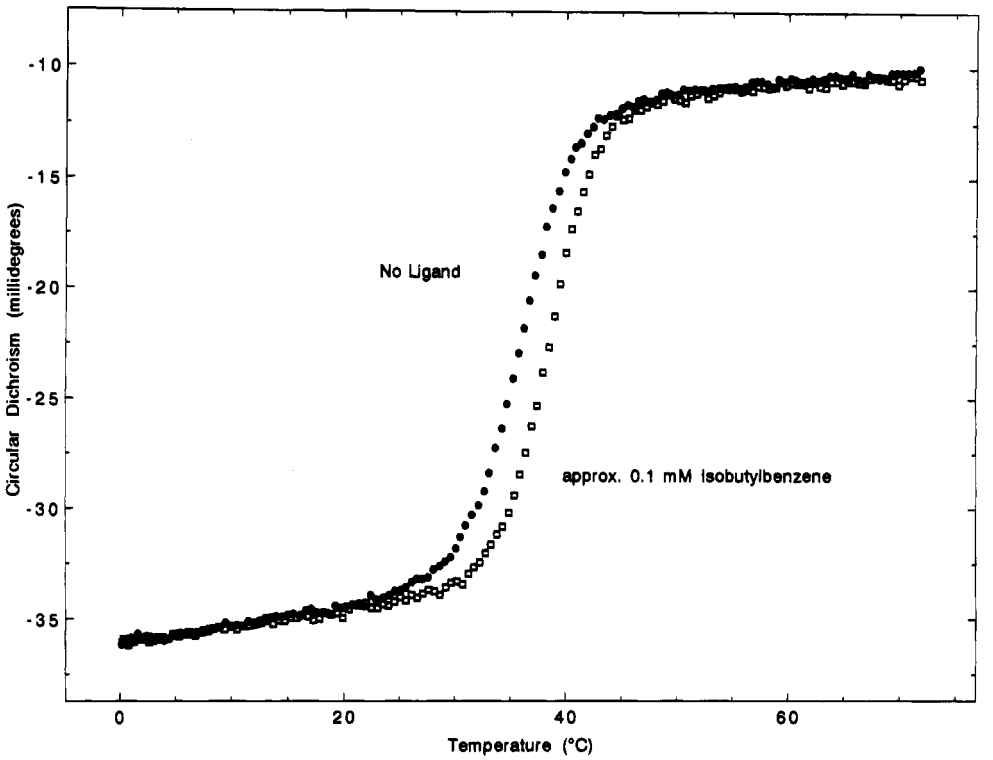

FIGURE 1: Representative thermal melting of mutant L99A of T4 lysozyme in the presence and absence of isobutylbenzene. The protein concentration is 1 μM in 25 mM KCl and 20 mM phosphate buffer, pH 3. T_m was 35.9 °C with ΔH at the T_m of 79.7 kcal/mol without ligand. T_m was 38.4 °C with ΔH at the T_m of 90.0 kcal/mol in about 0.1 mM isobutylbenzene.

EXPERIMENTAL PROCEDURES

The cysteine-free (Matsumura & Matthews, 1989) L99A lysozyme was isolated as described (Eriksson et al., 1993). Phenyl azide derivatives were the gift of Dr. J. Keana, University of Oregon. Other ligands were obtained from Aldrich and were typically reported as greater than 98% pure. Buffers were prepared using Baker analyzed reagents.

Initial Binding Screen. Binding of a ligand to the cavity-containing mutant was first tested using a "thermal upshift" assay (Eriksson et al., 1992a) in which the melting temperatures of the protein in the presence and absence of ligand were compared (Figure 1). Thermal unfolding was done in 0.025 M potassium chloride and 0.020 M potassium phosphate, pH 3.01, and monitored by circular dichroism (CD) at 223 nm (Eriksson et al., 1993). Under these conditions, the melting temperature of the L99A protein is 36 °C. Typically, 0.1−3 μL of neat ligand was added to 3 mL of buffer and dissolved by vigorous shaking immediately prior to addition of protein. Thermal denaturation was monitored in a 1 cm path-length cell using an in-cell thermal probe and in-cell stirring.

Interaction of a ligand with a folded protein results in an increase in melting temperature of the protein in the presence of ligand (Schellman, 1976). Such increases were usually accompanied by an increase in the enthalpy of unfolding of the complex. An unchanged melting temperature in the presence of the compound indicates a lack of preferential interaction or failure to detect the interaction as a result of insufficient ligand solubility or too weak of an association. It is formally possible for a ligand to interact equally with both the native and denatured forms of the protein. It seemed unlikely that the number and strength of the binding sites would be equal in both forms. Thus most doubtful cases of binding were examined at two or three different free ligand levels. A decrease in melting temperature, which was observed for some compounds (phenol, for example), supports the idea of a binding interaction with the unfolded state and could mask binding to the L99A cavity.

Calorimetry. Quantitative estimates of association were obtained using the MC2 titration calorimeter from Microcal, Inc. The measurements were performed in 50 mM sodium acetate, pH 5.5, at 29 °C. Low association constants and low ligand solubility required that most titrations be performed by injecting protein into ligand, rather than the reverse. For benzene, ethylbenzene, and *p*-xylene, titrations were performed in both directions with good agreement. Control experiments with wild-type lysozyme showed no evidence for the binding of any ligands. This is consistent with crystallographic observations which also show no evidence for binding of benzene to wild-type lysozyme (Eriksson et al., 1992a) and no evidence for the binding of any of the other ligands to sites other than the cavity resulting from the L99A mutation (Morton & Matthews, 1995).

Data analysis was performed using the ORIGIN program (Microcal, Inc.), fitting the association constant, the molar enthalpy of the reaction, and the ligand concentration to a single-site binding isotherm (Wiseman et al., 1989). It proved difficult to maintain aqueous solutions of volatile ligands at predetermined concentrations. Since evidence from calorimetry, crystallography (Morton & Matthews, 1995), and NMR (V. Feher and F. W. Dahlquist, personal communication) suggests that the stoichiometry of binding is 1:1, the concentration of ligand in the stock solution rather than the reaction stoichiometry was treated as a variable. In the case where the ligand is the titrand, these two variables are mathematically equivalent. Binding free energies were calculated as $-RT \ln K$, using a 1 M standard state. Solvent-transfer free energies for transfer from water to octanol at

Table 1: Binding of Ligands to L99A Lysozyme[a]

ligand	formula	mol wt	ΔT_m (°C)	ligand	formula	mol wt	ΔT_m (°C)
(a) ligands for which binding was detected				(b) ligands for which binding was not detected			
bromoethane	CH_3CH_2Br	108.97	2.4	chloroform	$CHCl_3$	119.38	
iodoethane	CH_3CH_2I	155.97	1.6	methanol	CH_3OH	32.00	
2-iodopropane	$(CH_3)_2CHI$	169.99	1.1	2-bromoethanol	$BrCH_2CH_2OH$	124.97	
ethyl disulfide	$(CH_3CH_2S)_2$	122.25	3.5	2-iodoethanol	ICH_2CH_2OH	171.97	
allyl ethyl sulfoxide	$H_2C{=}CHCH_2SO_2CH_2CH_3$	134.18	1.9	ethanol	C_2H_5OH	46.07	
allyl ethyl sulfide	$H_2C{=}CHCH_2SCH_2CH_3$	102.18	1.9	methyl sulfoxide	$(CH_3)_2SO$	78.13	
hexafluorobenzene	C_6F_6	186.05	1.5	1-propanol	$CH_3(CH_2)_2OH$	60.10	
fluorobenzene	C_6H_5F	96.10	5.9	furan	C_4H_4O	68.08	
iodobenzene	C_6H_5I	204.01	2.6	2-methylpropane at 2 atm	$CH(CH_3)_3$	56.12	
phenyl azide	$C_6H_5N_3$	119.14	1.6	pyridine	C_5H_5N	79.10	
nitrobenzene	$C_6H_5NO_2$	123.11	0.7	1,1-diethylurea	$(C_2H_5)_2NCONH_2$	116.16	
benzene	C_6H_6	78.10	5.9	2,3,4,5,6-pentafluoro-phenyl azide	$C_6F_5N_3$	209.08	
2-ethylfuran	$CH_3CH_2C_4H_3O$	96.13	1.9	2,3,4,5,6-pentafluoro-aniline	$C_6F_5NH_2$	183.08	
cyclohexene	C_6H_{10}	82.15	1.0	2,6-difluorophenyl azide	$F_2C_6H_3N_3$	155.12	
allyl sulfide	$(H_2C{=}CHCH_2)_2S$	114.21	1.6	1,2-diiodobenzene	$C_6H_4I_2$	329.91	
allyl propyl ether	$H_2C{=}CHCH_2O(CH_2)_2CH_3$	100.16	1.4	1,4-diiodobenzene	$C_6H_4I_2$	329.91	
isopropyl disulfide	$((CH_3)_2CHS)_2$	150.31	0.7	phenol	C_6H_5OH	94.11	
propyl disulfide	$(CH_3CH_2CH_2S)_2$	150.31	2.0	aniline	$C_6H_5NH_2$	93.13	
1,2-benzisoxazole	C_6H_4CHNO	119.12	0.4	2,5-dimethylfuran	C_6H_8O	96.13	
o-iodotoluene	$CH_3C_6H_4I$	218.04	1.9	cyclohexane	C_6H_{12}	84.16	
m-iodotoluene	$CH_3C_6H_4I$	218.04	2.0	2,1-benzisoxazole	C_7H_5NO	119.12	
p-iodotoluene	$CH_3C_6H_4I$	218.04	1.3	benzaldehyde	C_6H_5COH	106.12	
p-methylphenyl azide	$CH_3C_6H_4N_3$	133.16	1.4	benzoic acid	C_6H_5COOH	122.12	
toluene	$C_6H_5CH_3$	92.14	4.4	*syn*-benzaldehyde oxime	$C_6H_5CH{=}NOH$	121.14	
anisole	$C_6H_5OCH_3$	108.14	1.4	benzyl alcohol	$C_6H_5CH_2OH$	108.14	
benzyl mercaptan	$C_6H_5CH_2SH$	124.21	2.1	*p*-cresol	$CH_3C_6H_4OH$	108.14	
thioanisole	$C_6H_5SCH_3$	124.21	1.4	1-heptanol	$CH_3(CH_2)_6OH$	116.20	
methylcyclohexane	$C_6H_{11}CH_3$	98.19	1.1	*N,N'*-dimethylaniline	$C_6H_5N(CH_3)_2$	121.18	
phenylacetylene	$C_6H_5C{\equiv}CH$	120.14	0.7	1-octanol	$CH_3(CH_2)_7OH$	130.23	
2,3-benzofuran	$C_6H_4(CH)_2O$	118.14	3.0	butyl disulfide	$(C_4H_9S)_2$	178.36	
thianaphthene	$C_6H_4(CH)_2S$	134.20	1.5	mesitylene	$C_6H_3(CH_3)_3$	120.19	
indole	$C_6H_4(CH)_2NH$	117.15	1.7	quinoline	C_9H_7N	129.16	
styrene	$C_6H_5CH{=}CH_2$	104.15	2.0	*trans*-cinnamaldehyde	$C_6H_5CH{=}CHCHO$	132.16	
ethylbenzene	$C_6H_5CH_2CH_3$	106.17	3.8	azulene	$C_{10}H_8$	128.17	
o-xylene	$C_6H_4CH3)_2$	106.17	2.9	*tert*-butylbenzene	$C_6H_5C(CH_3)_3$	134.22	
m-xylene	$C_6H_4(CH3)_2$	106.17	1.8	(+)-camphene	$C_{10}H_{16}$	136.24	
p-xylene	$CH_3C_6H_4CH_3$	106.17	2.4	(−)-camphene	$C_{10}H_{16}$	136.24	
1-phenylpropyne	$C_6H_5C{\equiv}CCH_3$	116.16	3.0	(±)-camphor	$C_{10}H_{16}O$	152.24	
indene	$C_6H_4(CH)_2CH_2$	116.16	1.2	1-phenylheptane	$C_6H_5(CH_2)_6CH_3$	176.30	
2-methylbenzofuran	C_9H_8O	132.16	1.1				
n-propylbenzene	$C_6H_5(CH_2)_2CH_3$	120.20	6.2				
o-ethyltoluene	$CH_3CH_2C_6H_4CH_3$	120.20	1.1				
m-ethyltoluene	$CH_3CH_2C_6H_4CH_3$	120.20	1.1				
p-ethyltoluene	$CH_3CH_2C_6H_4CH_3$	120.20	1.7				
3-phenyl-1-propanol	$C_6H_5(CH_2)_3OH$	136.19	1.0				
3-phenylpropyl mercaptan	$C_6H_5(CH_2)_3SH$	152.26	4.5				
naphthalene	$C_{10}H_8$	128.18	1.0				
n-buytlbenzene	$C_6H_5(CH_2)_3CH_3$	134.22	1.4				
(±)-*sec*-butylbenzene	$C_6H_5CHCH_3C_2H_5$	134.22	1.3				
isobutylbenzene	$C_6H_5CH_2CH(CH_3)_2$	134.22	2.9				
amylbenzene	$C_6H_5(CH_2)_4CH_3$	148.25	1.2				
1-phenylhexane	$C_6H_5(CH_2)_5CH_3$	162.30	0.5				

[a] This table lists those compounds for which binding to the L99A cavity of T4 lysozyme was screened by means of a "thermal upshift" assay. ΔT_m gives the maximal increase in the melting temperature that was observed. Estimation of the lower limits of detection of this assay is difficult since optical transparency, compound solubility, and weak binding, as well as the error limits for ΔT_m, can all limit the detection of binding.

room temperature were taken from Sangster (1989) while those for transfer from water to the vapor phase were taken from Hine and Mookerjee (1975). In both cases, the transfer free energies are based on a 1 M standard state.

RESULTS

Initial Binding Screen. The results of the initial screen for ligand binding, based on the effect on the thermal unfolding transition, are presented in Table 1. Several trends are apparent. First, binding was not detected for any compound, such as aniline, pyridine, and quinoline, that carries a net charge at pH 3. Second, binding was not detected for most polar compounds, including ethanol, furan, phenol, and benzyl alcohol. Binding was, however, detected for indole and 3-phenyl-1-propanol. Third, binding was not detected for *tert*-butylbenzene, (±)-camphene, and (±)-camphor. This suggests that such molecules may be too large to fit inside the cavity, although binding of the *n*, *sec*, and *iso* isomers of butylbenzene was detected in the assay.

Energetics of Binding. Three subsets of those ligands shown to bind by the thermal upshift assay were chosen for quantitative analysis (Figure 2). Class I, the "isophobics",

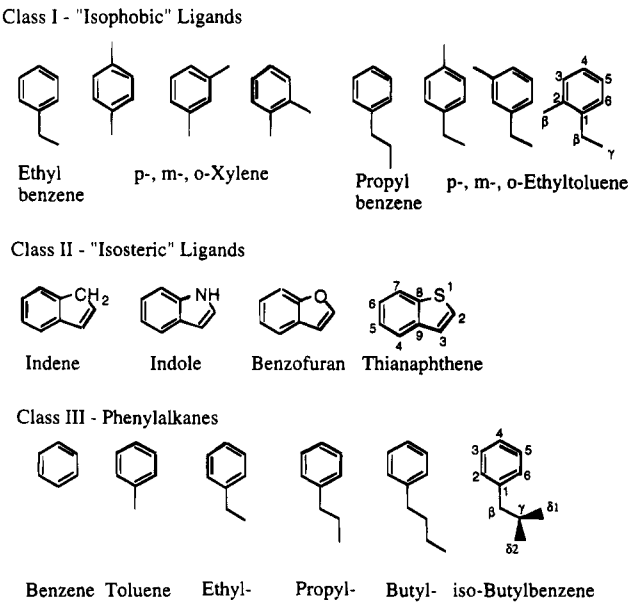

FIGURE 2: Three classes of ligands used for analysis of cavity binding. For reference, one ligand from each class is shown with atom labels. Substituent atoms are labeled α, β, etc. by analogy with protein side chains.

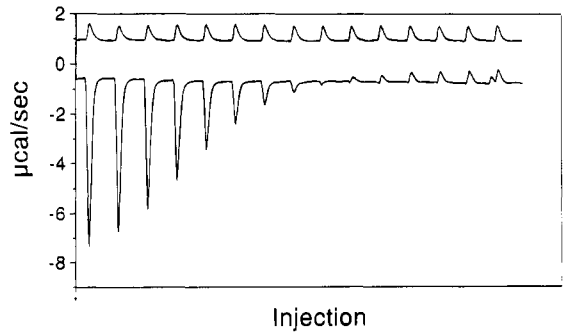

FIGURE 3: Representative titration profile for isobutylbenzene (~0.1 mM) titrated with L99A (4 mM) in 50 mM sodium acetate, pH 5.5. The offset upper trace shows L99A titrated into buffer without ligand. Injections of 10 μL of the protein solution were made every 2.5 min into the 1.4 mL reaction cell. After subtraction of blank runs, titrations were fit as described under Experimental Procedures to obtain the data in Table 2.

includes structural isomers of ethylbenzene and propylbenzene. These isomers have transfer free energies similar to their respective parents and were chosen to probe various regions of the cavity wall for possible differences in steric constraints in the presence of equivalent hydrophobic effects. Class II, the "isosterics", contains four isosteric molecules of varying hydrophobicity. Because these molecules are expected to have similar steric interactions with the cavity, they should, in principle, allow a direct analysis of the hydrophobic contribution to the binding energy. Class III is comprised of monosubstituted alkylbenzenes of increasing side-chain length. These were chosen to assess the relation between ligand size and binding energy in a manner similar to the protein mutagenesis studies of the type Gly → Ala → Val → Leu, etc.

Titration calorimetry was used to determine quantitative association constants at 29 °C for the binding to L99A of the ligands described above. A representative titration is shown in Figure 3. The binding energetics are presented in Table 2. Dissociation constants for the various ligands range from 14 to 500 μM, comparable to many enzyme–substrate

Table 2: Calorimetric Analysis of Ligand Binding[a]

ligand	$K_a \times 10^{-3}$ (M^{-1})	$-\Delta H$ (kcal/mol)	$-RT \ln K_a$ (kcal/mol)
benzene	5.7 ± 1.7	6.32 ± 0.37	−5.19 ± 0.16
ethylbenzene	14.8 ± 1.7	6.76 ± 0.87	−5.76 ± 0.07
o-xylene	2.13 ± 0.22	8.45 ± 0.96	−4.6 ± 0.06
m-xylene	2.75 ± 0.8	6.04 ± 0.03	−4.75 ± 0.15
p-xylene	2.37 ± 0.25	6.97 ± 0.98	−4.67 ± 0.06
propylbenzene	55.2 ± 2.0	9.97 ± 0.05	−6.55 ± 0.02
2-ethyltoluene	1.98 ± 0.20	7.71 ± 0.74	−4.56 ± 0.06
3-ethyltoluene	5.05 ± 0.15	7.84 ± 0.02	−5.12 ± 0.02
4-ethyltoluene	8.33 ± 0.08	8.44 ± 0.03	−5.42 ± 0.01
benzofuran	8.9 ± 0.5	8.04 ± 0.44	−5.46 ± 0.03
indene	5.17 ± 0.09	8.31 ± 0.48	−5.13 ± 0.01
indole	3.45 ± 0.38	11.23 ± 0.94	−4.89 ± 0.06
thianaphthene	13.6 ± 1.2	7.03 ± 0.04	−5.71 ± 0.05
toluene	9.8 ± 0.6	6.53 ± 0.73	−5.52 ± 0.04
isobutylbenzene	51.0 ± 4.9	7.09 ± 0.35	−6.51 ± 0.06
n-butylbenzene	69.8 ± 2.9	8.06 ± 0.98	−6.7 ± 0.02

[a] K_a is the association constant and ΔH the molar enthalpy of binding of the ligand to L99A lysozyme. Errors are given as the standard deviation of the mean calculated from multiple runs except in the cases of *m*-xylene, 2-ethyltoluene, propylbenzene, and thianaphthene, where the errors given are based on the goodness of the fit to the data.

binding constants. For all ligands, the molar enthalpy of binding is large and negative, unlike the enthalpy of transfer of liquid hydrocarbons from water to the neat organic phase, which is typically very close to zero at room temperature (Gill et al., 1976).

The observed binding energies of the ligands are compared with their free energies of transfer from water to vapor in Figure 4 and from water to octanol in Figure 5. Among the class I molecules, the xylenes and ethyltoluenes bind much more poorly than ethyl- and propylbenzene, respectively, and their binding free energies do not correlate well with transfer free energies (Figures 4a, 5a). Furthermore, there is no agreement between the binding free energies of the different structural isomers of ethylbenzene and those of propylbenzene. This indicates that the binding energetics of the class I molecules are strongly influenced by steric factors.

The class II molecules show only a rough correlation between their binding and transfer free energy, despite their nearly identical shapes and sizes (Figures 4b, 5b).

Among the class III molecules, binding becomes tighter as the side-chain becomes longer, up to a maximum of four carbons. This is in accord with the expectation that the hydrophobic effect provides a large contribution to binding free energy. Figure 5c shows the relation between binding energy and free energy of solvation as reflected in water–octanol partition coefficients (slope = 0.56; $R = 0.97$).

Entropic Consequences of Binding. To directly compare the observed binding free energies with solvent-transfer free energies, we must account for those contributions to the binding free energy that arise from purely statistical sources and which differ between the binding and solvent-transfer processes. One such contribution arises from the entropic cost of constraining a ligand to occupy a single conformation in the binding site, relative to its translational, rotational, and internal degrees of freedom in solution. In the present analysis, we assume that the ligands lose all rotational and translational freedom in the bound state and that their vibrational partition function does not change upon binding. For each such mode the cost of constraining the ligand is given by $\Delta G \simeq \Delta A = -RT \ln q$, where q is the partition

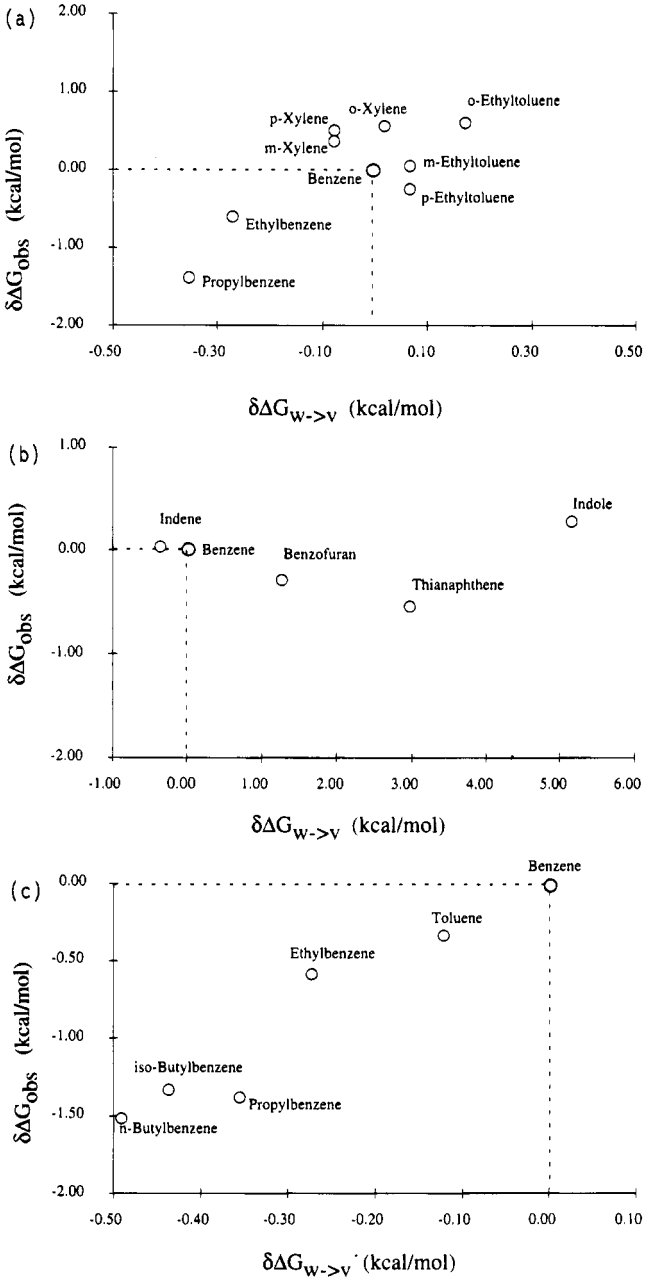

FIGURE 4: Binding energy relative to benzene ($\delta\Delta G_{OBS}$) for the three classes of ligands plotted *versus* free energy of transfer from water to vapor, relative to benzene ($\delta\Delta G_{W\rightarrow V}$). The symbol δ is used to denote that the values plotted are the difference between the value for the ligand and that for benzene. Data are taken from Table 5A.

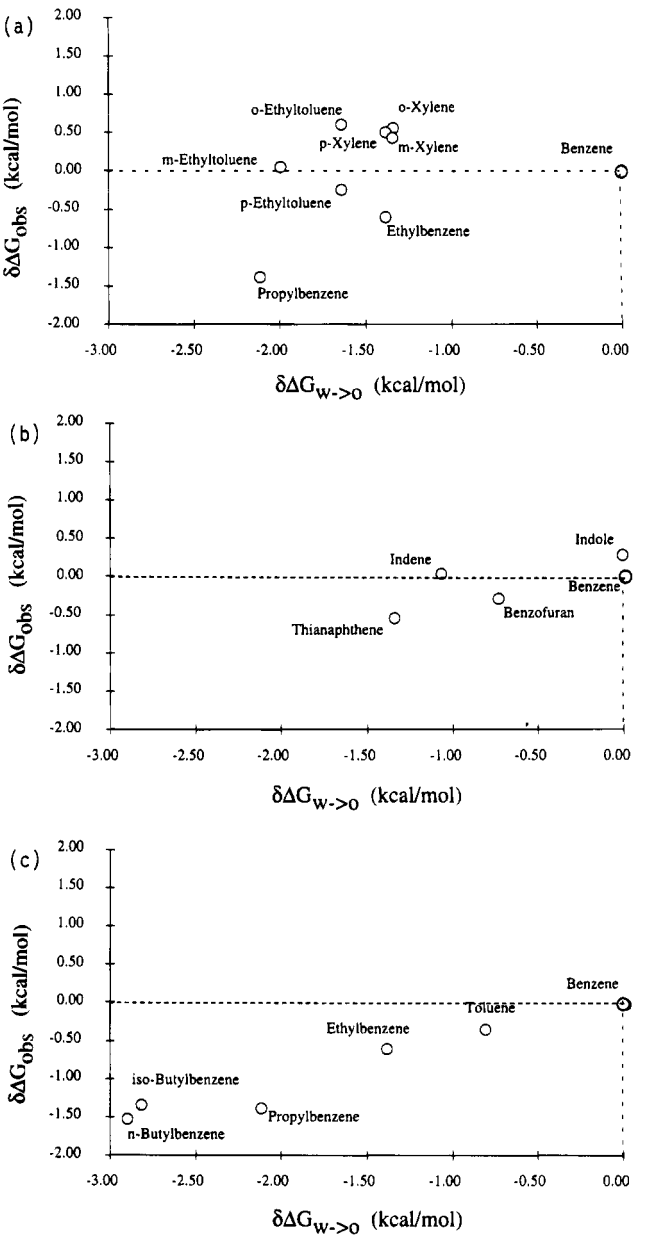

FIGURE 5: Binding energy, relative to benzene, for the three classes of ligands plotted *versus* transfer free energy from water to octanol, also relative to benzene. Values of $\delta\Delta G_{OBS}$ and $\delta\Delta G_{W\rightarrow O}$ are from Table 5B.

function for that particular mode. For simplicity, we consider only the change in free energy relative to that for benzene and denote such *relative* free energy differences using a small δ (e.g., $\delta\Delta G_{OBS}$ for ethylbenzene is the difference between ΔG_{OBS} for ethylbenzene and benzene).

The translational partition function is given (Hill, 1960; eq 9-1) by

$$q_T = V\left[\frac{2\pi kT}{h^2}m\right]^{3/2} \quad (1)$$

where V is the volume available to the ligand in solution, k and h are the Boltzmann and Planck constants, respectively, m is the mass of the ligand, and T is the temperature in kelvin. Since all of these values except the mass are the same for all the ligands, we have

$$q_T \propto m^{3/2}$$

Therefore, the differential effect on the binding energy, $\delta\Delta G_T$, for a ligand of mass m_2 relative to a ligand of mass m_1, due strictly to translational entropy, is given by

$$\delta\Delta G_T = +RT \ln [m_2/m_1]^{3/2} \quad (2)$$

In the case of butylbenzene relative to benzene this corresponds to $+RT \ln[(134/78)^{3/2}] = 0.48$ kcal/mol.

The rotational partition function of a polyatomic molecule is given (Hill, 1960; eq 9-8) by

$$q_R = (1/\sigma)\left[\frac{\pi T^3}{\theta_A\theta_B\theta_C}\right]^{1/2} \quad (3)$$

Table 3: Estimation of the Energy Cost of Localizing a Ligand on Binding, $\delta\Delta G_{STAT}$, Relative to That for Benzene[a]

ligand	mol wt	σ	n_{DIH}	inertia (D Å^2)$^{3/2}$	$\delta\Delta G_T$ (kcal/mol)	$\delta\Delta G_R$ (kcal/mol)	$\delta\Delta G_{DIH}$ (kcal/mol)	$\delta\Delta G_{STAT}$ (kcal/mol)
benzene	78	12	0	1175	0.00	0.00	0.0	0.00
ethylbenzene	106	1	0	3972	+0.27	+2.21	0.0	2.48
o-xylene	106	2	0	3814	+0.27	+1.77	0.0	2.04
m-xylene	106	2	0	4126	+0.27	+1.82	0.0	2.09
p-xylene	106	4	0	3818	+0.27	+1.36	0.0	1.63
propylbenzene	120	1	1	6674	+0.39	+2.52	0.4	3.30
2-ethyltoluene	120	1	1	5951	+0.39	+2.45	0.4	3.23
3-ethyltoluene	120	1	1	8770	+0.39	+2.68	0.4	3.46
4-ethyltoluene	120	1	0	6126	+0.39	+2.47	0.0	2.85
benzofuran	118	1	0	4032	+0.37	+2.22	0.0	2.59
indene	116	1	0	4384	+0.35	+2.27	0.0	2.62
indole	117	1	0	4126	+0.36	+2.23	0.0	2.59
thianaphthene	134	1	0	5912	+0.48	+2.44	0.0	2.93
toluene	92	2	0	1559	+0.15	+1.24	0.0	1.38
isobutylbenzene	134	1	1	8982	+0.48	+2.69	0.4	3.58
n-butylbenzene	134	1	2	1.08E+04	+0.48	+2.81	0.8	4.09

[a] σ is the internal point group symmetry number of the ligand; n_{DIH} is the number of σ bonds which become frozen upon binding; inertia is the square root of the product of the three principal moments of inertia of the ligands, calculated using EDPDB (Zhang & Matthews, 1995); $\delta\Delta G_T$, the change in translational entropy, is calculated according to eq 2; $\delta\Delta G_R$, the change in rotational entropy, is calculated according to eq 4; $\delta\Delta G_{DIH}$, the change in free energy due to restriction of dihedral rotation upon binding, is calculated from eq 5; $\delta\Delta G_{STAT}$ is the total conformational and configurational effects of immobilizing the ligand in the binding site, relative to benzene.

where θ_A, for example, equals $h^2/8\pi^2kI_A$ and I_A, I_B, and I_C are the principal moments of inertia. σ is the point group symmetry number of the ligand. We can again discard the constant terms, leaving

$$q_R \propto \left[\frac{I_A I_B I_C}{\sigma^2}\right]^{1/2}$$

The principal moments of inertia were calculated from the atomic coordinates of the ligands. While it is true that the conformations of some ligands and, concomitantly, their moments of inertia can change upon binding, these effects are small relative to the total correction being applied.

The differential binding free energy for molecule *j* relative to *i* that arises from the rotational partition function is thus given by

$$\delta\Delta G_R = +RT \ln\left[\frac{(I_{A,j}I_{B,j}I_{C,j})^{1/2}\sigma_i^2}{(I_{A,i}I_{B,i}I_{C,i})^{1/2}\sigma_j^2}\right] \quad (4)$$

For butylbenzene relative to benzene, this corresponds to $+RT \ln[(10840 \times 12)/(1175 \times 1)] = +2.82$ kcal/mol. This difference arises from both their principal moments of inertia and their different symmetry numbers.

Molecules for which internal rotation is restricted upon binding lose additional entropy. The amount of entropy lost upon fixing a rotatable dihedral bond in long-chain alkanes has been estimated as roughly 1.5 cal/(mol·K) per dihedral bond (Mark & Tobolsky, 1950). This estimate is in good agreement with more recent estimates of the cost of constraining amino acid side-chain conformation during protein folding (Pickett & Sternberg, 1993; Creamer & Rose, 1992). In terms of free energy, the cost is given by

$$\Delta G_{DIH} = n_{DIH} T\delta S_{DIH} = n_{DIH}(0.45 \text{ kcal/mol}) \quad (5)$$

where n_{DIH} is the number of dihedral bonds that become frozen upon binding.

A terminal methyl group was assumed to be equally unconstrained in the bound and free states, and no correction was made. In the case of monosubstituted phenyl rings, such as monoalkylbenzenes, the bond that connects the ring to the first methylene has two internal energy minima. Where these two are indistinguishable, due to local 2-fold symmetry, binding does not alter the entropy of the bond and no correction need be applied.

The sum of all these statistical corrections, ΔG_{STAT}, is given by

$$\delta\Delta G_{STAT} = \delta\Delta G_T + \delta\Delta G_R + \delta\Delta G_{DIH} \quad (6)$$

Values of $\delta\Delta G_{STAT}$ are collected in Table 3.

Scaled Particle Theory Calculations. It is also necessary to obtain an estimate of the cost of forming a cavity in octanol sufficiently large to accommodate a given ligand (ΔG_{CAV}). This can be obtained using scaled particle theory (Reiss, 1965; Lee, 1991). The only information required is the density of the solvent and the molecular size of both the solvent and the solute cavity.

There are several ways to estimate molecular size, including calculation of the volume enclosed by the molecular surface (Richards, 1977; Connolly, 1983) and the group additivity approach of Bondi (1964). The latter has been shown to correspond closely to the value which gives the best fit between the observed thermodynamics of several real liquids and those predicted by scaled particle theory (Reiss, 1965), although volumes calculated in this way are typically smaller than those obtained from the molecular surface calculations (Table 4). For this reason, we chose the diameter of octanol to be equivalent to the diameter of a sphere having the same volume as the volume of octanol calculated according to Bondi. The actual calculations were performed according to eq 131 of Reiss (1965), encoded in the computer program SPT (H. Qian, unpublished). The results of these calculations are presented in Table 4.

DISCUSSION

Many studies have shown that the substitution of one type of amino acid for another in the core of a protein can have different consequences for protein stability, depending on the particular site of substitution [e.g., Shortle et al. (1990) and Eriksson et al. (1992b)]. This is because in addition to

Table 4: Energy Cost of the Formation of a Cavity in Octanol Sufficient To Accommodate a Given Ligand[a]

	Connolly volume			Bondi volume		cavity cost	
ligand	area (Å^2)	volume (Å^3)	equivalent radius (Å)	volume (Å^3)	equivalent radius (Å)	Connolly (kcal/mol)	Bondi (kcal/mol)
benzene	96	77	2.64	80	2.68	7.68	7.87
ethylbenzene	136	120	3.06	116	3.02	9.77	9.56
o-xylene	138	123	3.09	122	3.07	9.92	9.82
m-xylene	140	124	3.10	122	3.07	9.98	9.82
p-xylene	141	122	3.08	122	3.07	9.87	9.82
propylbenzene	150	141	3.23	133	3.17	10.68	10.35
2-ethyltoluene	141	127	3.12	139	3.21	10.08	10.57
3-ethyltoluene	159	144	3.25	139	3.21	10.79	10.57
4-ethyltoluene	158	143	3.25	139	3.21	10.79	10.57
benzofuran	125	107	2.95	106	2.94	9.20	9.14
indene	134	113	3.00	114	3.01	9.45	9.50
indole	127	109	2.96	111	2.98	9.25	9.35
thianaphthene	139	128	3.13	115	3.02	10.12	9.56
toluene	111	92	2.80	99	2.87	8.45	8.79
isobutylbenzne	161	154	3.33	150	3.29	11.24	11.02
n-butylbenzene	167	161	3.37	150	3.29	11.47	11.02

[a] The surface area and volume of each ligand were calculated with the program MS (Connolly, 1983), using a probe of radius 1.9 Å, corresponding to the radius of a methyl group. Volumes were also calculated according to Bondi (1964). In each case, the "equivalent radius" is the radius of a sphere of equal volume. The cost of forming a cavity in octanol was calculated using eq 131 of Reiss (1965) with the hard sphere diameter of octanol taken to be 6.66 Å.

the hydrophobic effect context-dependent steric interactions and entropic effects are also important. The contributions of effects other than hydrophobicity have also been seen in studies of protein−ligand interactions (Estell et al., 1986; Bigler et al., 1993). The work described here attempts to experimentally separate these effects from one another and from other factors such as hydrogen bonding and salt bridge formation. In particular, we try to estimate the energetic contributions arising from packing effects and from configurational and conformational statistics relative to the contribution from solvent transfer.

The cavity formed by the L99A mutation in the core of T4 lysozyme provides an optimal binding site for such an experiment. It is shielded from bulk solvent, simplifying estimation of desolvation effects upon binding. It is nonpolar, with no unsatisfied hydrogen bond donors or acceptors and no formal charges. [In principle, the carbonyl oxygens which are involved in intrahelical hydrogen bonds can accept additional hydrogen bond donors as can the S^{δ} atom of Met 102, which is unpaired in the L99A apoprotein structure (Morton & Matthews, 1995).] The aromatic ring of Phe 153 is within the wall of the cavity but is not positioned so as to allow specific aromatic−aromatic interactions with bound ligands (Burley & Petsko, 1986). These features limit the types of interaction that are possible between the protein and the ligand.

The initial binding screen (Table 1) clearly demonstrates that hydrophobicity is an important component of the binding free energy. In general, ligands with polar groups or formal charges do not bind. Two clear-cut examples are provided by bromoethanol *versus* bromoethane and by iodoethanol *versus* iodoethane. These compounds are all small enough to be readily accommodated within the cavity but differ substantially in polarity. The thermal upshift assay indicates that only the nonpolar compounds, bromoethane and iodoethane, bind. There is no evidence for the binding of either bromoethanol or iodoethanol. (The use of these compounds was suggested to us by W. W. Cleland as controls to test whether small, relatively polar, compounds bind in the cavity.)

The thermal upshift screen also reveals that changes in ligand size or shape can have large effects on binding. A particularly clear example is provided by the isomeric molecules *n*- and *tert*-butylbenzene, one of which binds and the other does not. These results indicate that both steric and hydrophobic effects influence binding.

Binding Energetics. Perhaps the most striking aspect of the energetics of binding (Table 2) is the uniformly high values of the binding enthalpies. Thermodynamic hallmarks of the hydrophobic effect at room temperature include large ΔC_p and small ΔH: the room temperature transfer of nonpolar molecules from water to nonpolar phases is largely entropy-driven. Ligand binding in the present system, however, is largely enthalpy-driven. Benzene, for example, has a binding enthalpy at 29 °C of −6.3 kcal/mol but a binding free energy of only −5.2 kcal/mol. For comparison, its enthalpy of transfer from the neat organic phase to water is −0.72 kcal/mol at 29 °C (Gill et al., 1975).

These results indicate that the binding free energies of the various ligands are dominated by enthalpy, not by entropy, as might have been expected on the basis of the thermodynamics of solvent transfer. Presumably, the relatively large enthalpic component arises from interactions between the protein and ligand that are not represented in the solvent-transfer process. In light of this result, it is difficult to describe the association of the ligands with the protein strictly in terms of aqueous−organic solvent transfer.

Specific Binding Models. The observed binding reaction may be analyzed by dividing it into distinct processes. There are many ways to make these divisions; we consider two, shown as Scheme 1 and Scheme 2, in Figure 6. [For a similar analysis of the effects of the cavity-creating mutation itself, see Lee (1993).] Both schemes consider binding as composed of three steps: (1) desolvation of the ligand and binding site; (2) immobilization of the ligand relative to the binding site; (3) reorganization of the binding site to its final conformation and introduction of the intermolecular potential between the protein and the ligand. This last step corresponds to the packing of the ligand in the specific environment of the binding site. By estimating the free energy

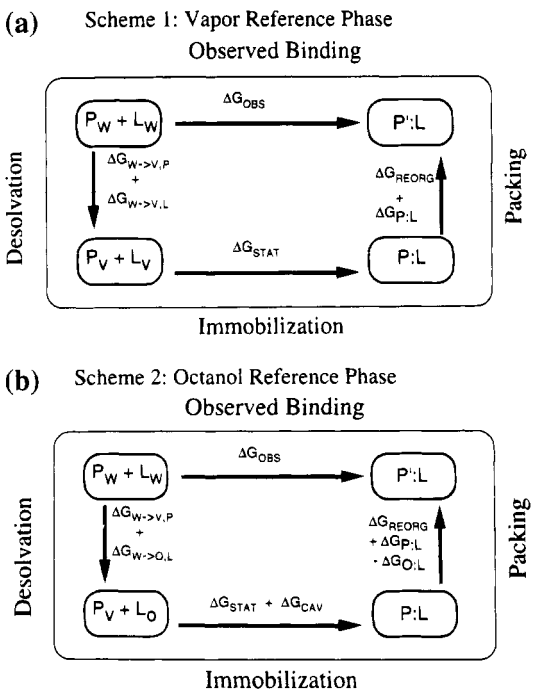

FIGURE 6: Thermodynamic cycles representing two possible ways to describe the binding reactions. In each case the observed, overall reaction is indicated by the top side of the cycle. The other three sides represent desolvation, immobilization, and packing processes, respectively. Subscripts W, V, and O represent aqueous, vapor, and organic phases, respectively. L represents the ligand, and P represents the internal binding site of the L99A protein (as opposed to the entire protein). (a) In the vapor scheme (1), desolvation is modeled as the cost of transferring both the binding site and the ligand from aqueous solution to the vapor phase ($\Delta G_{W\rightarrow V}$). The second step corresponds to the immobilization of the ligand in the binding site (ΔG_{STAT}) *without* introducing any ligand−protein potential. In the third step, the protein is forced to adopt the conformation seen in the complex (ΔG_{REORG}), followed by introduction of the protein−ligand interactions ($\Delta G_{P:L}$). The reorganization of the protein is indicated by a prime. (b) In the octanol scheme (2), desolvation of the ligand is modeled as transfer from water into octanol ($\Delta G_{W\rightarrow O}$). The second step again corresponds to transferring the ligand into the binding site (ΔG_{STAT}), but in this case the work available from closing the cavity remaining in the octanol solution must also be included (ΔG_{CAV}). (All other cavity openings and closings are included in the respective desolvation steps and need not be considered explicitly.) In the third step, the protein is forced to reorganize to accommodate the ligand (ΔG_{REORG}). In this case, the protein−ligand interactions are introduced ($\Delta G_{P:L}$) and the octanol−ligand interactions (deferred from the preceding step; see text) are removed ($-\Delta G_{O:L}$). Deferring this last process until step 3 allows the total residual packing energy to serve as a guide to the octanol-like character of the protein interior. Use of these binding cycles requires the assumption that the various free energy terms describe independent transformations between states. This seems particularly reasonable in the desolvation and statistical steps. This independence is less obvious among the terms in the packing step. It is less important there, however, because the packing free energies are obtained as residual free energies and are not further separated into ΔG_{REORG}, $\Delta G_{P:L}$ and $\Delta G_{O:L}$.

associated with the first two steps, we are able, by subtraction, to infer the packing free energy. The two schemes differ in the choice of the nonpolar phase in step 1 and in the consequences of that choice for the remaining steps. In particular, the packing terms derived from the two different analyses correspond to two different processes and provide two views of the role played by the steric interactions between the ligand and the binding site.

Step 1. In Scheme 1 (Figure 6a) the desolvation of the ligand and the binding site are modeled by the transfer of each from aqueous solution into the vapor phase. For the ligand, this corresponds to the process of vaporization from dilute aqueous solution. The free energy can be obtained from the aqueous solubility and the equilibrium vapor pressure of the ligand (Hine & Mookerjee, 1975; TRC Tables, 1986). For the binding site, the desolvation corresponds to the removal of any solvent that may be present in the apoprotein, leaving the cavity empty and keeping it in the conformation seen in the apoprotein. The free energy associated with this process is difficult to estimate, although Wolfenden and Radzicka (1994) have argued that cavities like that created by the L99A mutation remain empty most of the time. Regardless of the exact cost of desolvating the binding site, however, this cost will be the same for all the ligands. The relative cost of the total desolvation for association of any ligand relative to benzene ($\delta\Delta G_{W\rightarrow V}$, where the subscript W→V indicates transfer from water to vapor) is then simply the difference between the vaporization free energies of the two ligands (i.e., $\Delta G_{W\rightarrow V,LIGAND} - \Delta G_{W\rightarrow V,BENZENE}$).

Scheme 2 (Figure 6b) differs from Scheme 1 in that the ligand is transferred not to vapor but to octanol. The energetics of this transfer are very different from the transfer to the vapor phase for two reasons. First, a ligand-sized cavity must be opened in the octanol to accommodate the ligand. Second, the ligand makes significant, favorable interactions with the octanol, whereas none are made in the vapor phase. While the effects of these differences are included in the measured transfer free energies, and need not be considered explicitly during step 1, they lead to differences between the two schemes in subsequent steps. Scheme 2 has the advantage of treating the desolvation component of binding as a transfer from water to a nonpolar solvent and thus more faithfully reflects the standard understanding of the hydrophobic effect. As with Scheme 1, the cost of desolvating the binding site may be circumvented by considering ligand-transfer free energies relative to benzene ($\delta\Delta G_{W\rightarrow O}$, where the subscript w→o indicates transfer from water to octanol).

Step 2. In both schemes, the second step includes the entropic cost of immobilizing the ligand as it binds, which can be quite substantial (Table 3) [see also Janin and Finkelstein (1990)]. The second step is more complicated in Scheme 2 than in Scheme 1 because Scheme 2 requires that the octanol cavity left by the ligand upon transfer to the protein binding site be closed. The calculation of the work involved in this process by means of scaled particle theory is sensitive to the exact values of the molecular volumes used, and we therefore have included in Table 4 the results of alternative volume estimates based on Bondi (1964) and Connolly (1983). The disagreement is typically less than 0.4 kcal/mol. In addition to closing the cavity, the intermolecular potential between the ligand and the surrounding octanol must be removed. We choose to defer this removal until step 3.

Step 3. The third and final step involves the reorganization of the protein binding site to the conformation of the final complex (ΔG_{REORG}), followed by introduction of the protein−ligand interaction potential ($\Delta G_{P:L}$). The second scheme requires a third term ($-\Delta G_{O:L}$) to account for removal of the octanol−ligand interaction discussed above. In each of the two schemes the total free energy of step 3 is obtained by subtraction and is referred to as the packing free energy

($\Delta G_{PAK,V}$ for Scheme 1; $\Delta G_{PAK,O}$ for Scheme 2). In both schemes packing is defined, strictly speaking, only as a "residual" free energy: it is that part of the observed binding free energy that is not accounted for by the first two steps. In practice, however, we assign a molecular interpretation to this free energy. It corresponds to the work required to reorganize the binding site to accommodate the ligand, together with the steric interactions experienced by the ligand in the bound state relative to the desolvated state.

The most important and useful difference between Schemes 1 and 2 now becomes apparent. The reorganization part of step 3 is identical in both schemes, but the introduction of the interaction potential is not. In Scheme 1, this potential is the entire interaction energy between the ligand and the reorganized binding site; in Scheme 2, this potential corresponds to the *differential* interaction made between the ligand and the binding site, relative to that made between the ligand and octanol. The utility of this definition of packing is that it allows comparison between the ligand−octanol interactions and the ligand−protein interactions for each ligand and thus provides a means of assessing the usefulness of water−octanol transfer to describe water−protein transfer.

To summarize, the observed binding free energy, relative to benzene, may be expressed with reference to Scheme 1 as

$$\begin{aligned}\delta\Delta G_{OBS} &= \delta\Delta G_{W\rightarrow V} + \delta\Delta G_{STAT} + \\ &\qquad \delta\Delta G_{REORG} + \delta\Delta G_{P:L} \\ &= \delta\Delta G_{W\rightarrow V} + \delta\Delta G_{STAT} + \delta\Delta G_{PAK,V} \quad (7)\end{aligned}$$

For Scheme 2, the overall expression becomes

$$\begin{aligned}\delta\Delta G_{OBS} &= \delta\Delta G_{W\rightarrow O} + \delta\Delta G_{STAT} + \delta\Delta G_{CAV} + \\ &\qquad \delta\Delta G_{REORG} + (\delta\Delta G_{P:L} - \delta\Delta G_{O:L}) \\ &= \delta\Delta G_{W\rightarrow O} + \delta\Delta G_{STAT} + \delta\Delta G_{CAV} + \\ &\qquad \delta\Delta G_{PAK,O} \quad (8)\end{aligned}$$

The relative packing energies determined using these two schemes are given in Table 5.

Vapor-Phase Analysis. We consider first the calculations in which desolvation is modeled as transfer from water to the vapor phase (Scheme 1; Table 5A). The relations defined in Scheme 1 and eq 7 are clearly illustrated in Figure 7a−c. Among the aromatic hydrocarbons of classes I and III, the magnitude of the desolvation term, $\delta\Delta G_{W\rightarrow V}$, is typically smaller than the binding free energy, $\delta\Delta G_{OBS}$, suggesting that effects other than desolvation contribute favorably to binding. The free energy from statistical effects, however, is large and positive (unfavorable). The combination of small desolvation free energies and large statistical terms implies that the packing energies are large and favorable, as is seen (Table 5A; Figure 7a,b).

As discussed above, the packing energy consists of reorganization and interaction terms (ΔG_{REORG} and $\Delta G_{P:L}$ in Scheme 1). In absolute terms, the work required to reorganize the binding site from the apoprotein conformation to the conformation in the bound state is, by definition, always positive (unfavorable for binding), since conceptually the reorganization occurs prior to introduction of the protein−ligand potential. Because benzene is the smallest of the ligands studied here, and because the other ligands cause larger changes in the protein structure (Morton & Matthews, 1995), it is likely that the reorganization required for binding the other ligands is also unfavorable in a relative sense (i.e., $\delta\Delta G_{REORG} > 0$). The absolute interaction potential ($\delta\Delta G_{P:L}$) can be either positive or negative, depending on the geometry of the final complex. While it seems likely that the absolute interaction potential of benzene with the binding site is negative (i.e., favorable) the relative interaction potentials of the remaining ligands, relative to benzene, are

Table 5

(A) Contributions to Ligand Binding Based on Transfer to the Vapor Phase[a]

ligand	$\delta\Delta G_{OBS}$ (kcal/mol)	$\delta\Delta G_{W\rightarrow V}$ (kcal/mol)	$\delta\Delta G_{STAT}$ (kcal/mol)	$\delta\Delta G_{PAK,V}$ (kcal/mol)
class I				
ethylbenzene	−0.57	−0.27	2.48	−2.78
o-xylene	0.59	0.01	2.04	−1.47
m-xylene	0.44	−0.08	2.09	−1.57
p-xylene	0.53	−0.08	1.63	−1.02
propylbenzene	−1.36	−0.36	3.30	−4.31
2-ethyltoluene	0.63	0.17	3.23	−2.76
3-ethyltoluene	0.07	0.06	3.46	−3.45
4-ethyltoluene	−0.23	0.06	2.85	−3.14
class II				
benzofuran	−0.27	1.24	2.59	−4.09
indene	0.06	−0.40	2.62	−2.16
indole	0.30	5.13	2.59	−7.42
thianaphthene	−0.52	2.95	2.93	−6.40
class III				
benzene	0.00	0.00	0.00	0.00
toluene	−0.33	−0.12	1.38	−1.59
ethylbenzene	−0.57	−0.27	2.48	−2.78
propylbenzene	−1.36	−0.36	3.30	−4.31
isobutylbenzene	−1.32	−0.44	3.58	−4.45
n-butylbenzene	−1.50	−0.49	4.09	−5.10

(B) Contributions to Ligand Binding Based on Transfer to Octanol

ligand	$\delta\Delta G_{OBS}$ (kcal/mol)	$\delta\Delta G_{W\rightarrow O}$ (kcal/mol)	$\delta\Delta G_{STAT}$ (kcal/mol)	$\delta\Delta G_{CAV}$ (kcal/mol)	$\delta\Delta G_{PAK,O}$ (kcal/mol)
class I					
ethylbenzene	−0.57	−1.40	2.48	−1.69	0.03
o-xylene	0.59	−1.35	2.04	−1.95	1.85
m-xylene	0.44	−1.38	2.09	−1.95	1.68
p-xylene	0.53	−1.40	1.63	−1.95	2.24
propylbenzene	−1.36	−2.13	3.30	−2.48	−0.05
2-ethyltoluene	0.63	−1.65	3.23	−2.70	1.76
3-ethyltoluene	0.07	−2.01	3.46	−2.70	1.32
4-ethyltoluene	−0.23	−1.65	2.85	−2.70	1.27
class II					
benzofuran	−0.27	−0.74	2.59	−1.27	−0.84
indene	0.06	−1.08	2.62	−1.63	0.15
indole	0.30	−0.02	2.59	−1.48	−0.79
thianaphthene	−0.52	−1.35	2.93	−1.69	−0.42
class III					
benzene	0.00	0.00	0.00	0.00	0.00
toluene	−0.33	−0.82	1.38	−0.92	0.03
ethylbenzene	−0.57	−1.40	2.48	−1.69	0.03
propylbenzene	−1.36	−2.13	3.30	−2.48	−0.05
isobutylbenzene	−1.32	−2.83	3.58	−3.14	1.08
n-butylbenzene	−1.50	−2.91	4.09	−3.14	0.46

[a] Values of $\delta\Delta G_{OBS}$, the observed binding energy relative to benzene, are from Table 2. Values of $\delta\Delta G_{W\rightarrow V}$ correspond to the cost of transferring the ligand from dilute aqueous solution to the vapor phase, relative to the same cost for benzene, and are from Hine and Mookerjee (1975). Values of $\delta\Delta G_{STAT}$ are from Table 3. Values of $\delta\Delta G_{PAK,V}$ are calculated according to eq 7. [b] Values of $\delta\Delta G_{OBS}$, the observed binding energy relative to benzene, are from Table 2. Free energies of transfer from octanol to water at room temperature for the various ligands relative to benzene ($\delta\Delta G_{W\rightarrow O}$) are taken from Sangster (1989). $\delta\Delta G_{STAT}$ is defined as in (A). $\delta\Delta G_{CAV}$ corresponds to the work available from closing the cavity remaining in octanol following transfer of the ligand into the protein binding site, relative to the same value for benzene. $\delta\Delta G_{PAK,O}$ is the packing energy as defined by eq 8 in the text.

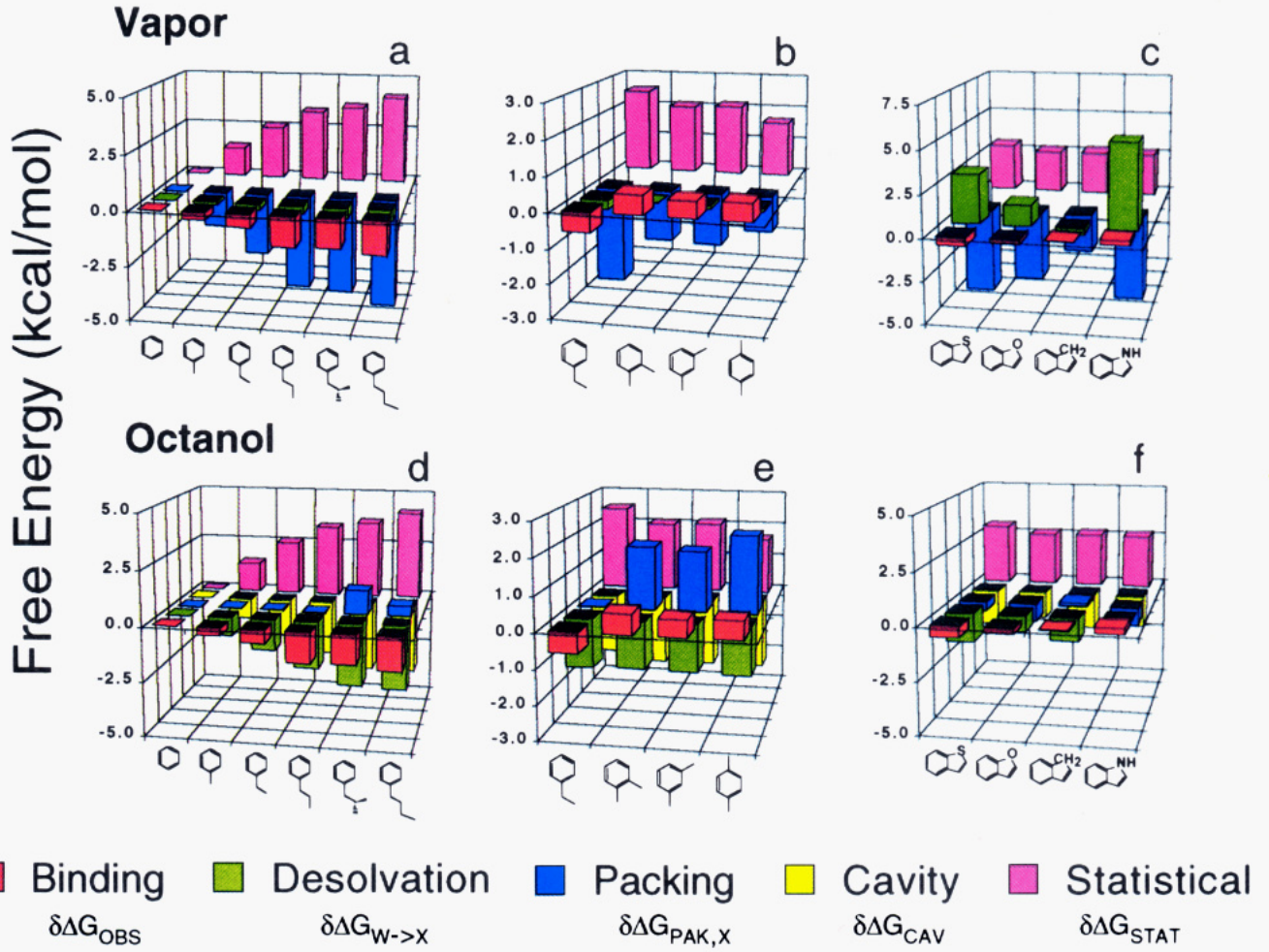

FIGURE 7: Bar graphs summarizing the factors that contribute to the binding of each ligand as estimated using Scheme 1 (vapor phase) and Scheme 2 (octanol phase), respectively, in Figure 6. Individual ligands are shown by their structures at the bottom of each panel. Panels a and d, b and e, and c and f summarize the analyses for class III, I, and II ligands, respectively. The identity of each term in the analysis is indicated at the bottom of the figure, where X represents either O or V (see text and legend to Figure 6 for details).

uncertain. That the relative packing energies of these ligands are negative despite a positive reorganization energy implies that their protein−ligand interaction potentials relative to benzene are in fact negative. In determining the packing energies of the class I and III ligands, the favorable interaction term outweighs the unfavorable reorganization term, even in the case of the largest ligands that bind, such as butylbenzene, or ligands such as xylenes with shapes different than the apoprotein binding site. The effect of the reorganization can nevertheless be seen in the details of the packing energies. For example, a single methylene increase from ethylbenzene to propylbenzene increases the stability of the complex by 1.5 kcal/mol, while further increments (which apparently require greater reorganization of the binding site) are less stabilizing (Table 5A).

In summary, the relative binding affinities of the aromatic hydrocarbons result from a compromise between large statistical forces that oppose binding and large packing forces that favor binding, with only slight modulation by the forces represented by desolvation from water to the vapor phase. This does not, of course, mean that the hydrophobic effect plays only a minor role in binding. In the first place, the values quoted in Table 5A are relative to benzene. In the second place, the hydrophobic effect, as traditionally understood, is not adequately represented by desolvation from water to vapor, partly because the solute cannot make intermolecular contacts in the vapor phase similar to those available in any condensed phase, be it aqueous solution, octanol, or the protein interior. Indeed, the utility of the water−vapor transfer free energies is that they allow assignment of an unambiguous physical meaning to the derived packing energies. This clear picture of packing, unfortunately, comes at the expense of foregoing the standard understanding of the hydrophobic effect, as reflected in the process of transfer between two condensed phases.

The relative binding energetics of the class II ligands cannot be simply characterized as the balance of two large opposing forces. This is because their free energies of transfer from water to vapor range quite widely, from −0.4 to 5.1 kcal/mol relative to the value for benzene (Table 5A; Figure 7c). This large range arises from two sources: the stronger electrostatic interactions with aqueous solvent made by the more polar ligands and the stronger effect of the less polar ligands on the structure of the aqueous solvent. In contrast to their water−vapor transfer free energies, the statistical free energies of the class II ligands are grouped closely because of their similar sizes and shapes. The packing energies for the class II ligands also display a wide range of values. One might assume that, because of their similar size and shape, these ligands would bind similarly and have similar reorganization energies, leaving the differences in their packing energies to be explained by their

interactions with the reorganized binding site. Crystallographic analysis does not bear this out, demonstrating instead significant differences in the degree of protein reorganization upon binding of the different ligands (Morton & Matthews, 1995). A more complete analysis of the interactions made by these ligands with the binding site is made possible by considering their packing energies derived from Scheme 2 (*vide infra*).

Octanol-Phase Analysis. Treating desolvation as transfer from water into octanol rather than into the vapor phase leads to several differences in the analysis of the binding energetics, as illustrated in Figure 7d–f. The relative differences in the water–octanol transfer free energies are generally much larger than their water–vapor counterparts. In most cases, they are in fact larger than the relative binding free energies (Table 5B; Figure 7d–f). Furthermore, the hypothetical transfer of the ligands from octanol into the protein binding site as part of the second step of binding requires that the cavity which remains in the octanol be closed. The requirement to explicitly include this step may be seen as a consequence of the fact that the binding site cavity already exists prior to binding. The corresponding free energy, $\delta\Delta G_{CAV}$, as estimated from scaled particle theory, is nearly as large as that associated with localizing the ligand ($\delta\Delta G_{STAT}$), but is of opposite sign. The estimated packing term ($\delta\Delta G_{PAK,O}$) corresponds to the sum of the cost of reorganization and the *differential* interactions made by the ligand with the protein relative to those made with octanol. Thus it is a crude measure of the difference between the environment provided by the protein interior and that corresponding to a ligand-sized cavity in octanol.

For the aromatic hydrocarbons of class III the values of $\delta\Delta G_{W\rightarrow O}$ are larger than the observed binding energies by a factor of roughly 2. There is also a fairly good correlation between the two (Figure 5c), despite the expectation that small differences in the geometry of the association might result in large differences in packing energy. Values of $\delta\Delta G_{STAT}$ are the same as for Scheme I. They are large but are not quite offset by $\delta\Delta G_{CAV}$, which is typically nearly as large as $\delta\Delta G_{STAT}$, but of opposite sign (Table 5B; Figure 7d). Also for the class III ligands, $\delta\Delta G_{PAK,O}$ is close to zero except for the largest compounds, isobutyl- and *n*-butylbenzene. The simplest interpretation is that both the reorganization and the relative interaction free energies are close to zero. This is consistent with the shape of the cavity as observed crystallographically, which shows room for both the aromatic ring and a single small substituent (Eriksson et al., 1992a; Morton & Matthews, 1995). The nonzero values for isobutyl- and *n*-butylbenzene reflect the higher cost of reorganizing the cavity to accommodate these larger ligands. If our interpretation is correct, the relative interactions between the class III ligands and the binding site, including any reorganization associated with binding, are quite well approximated by the relative interactions made between the class III ligands and octanol. For this group of ligands the protein provides a complementary binding site similar to that provided by an organic liquid. Despite this congruence, however, it is not appropriate to say that binding is dominated by hydrophobic effects. Both the work available from closing the octanol cavity and the cost of ligand immobilization are of greater magnitude than the solvent-transfer free energy (Table 5B; Figure 7d).

As expected, the water–octanol transfer free energies among the two sets of isomers that comprise class I are fairly similar. They are also significantly larger than the values of $\delta\Delta G_{OBS}$ (Figure 7e). As with the class III ligands, $\delta\Delta G_{STAT}$ and $\delta\Delta G_{CAV}$ are large and more or less compensating. The value of $\delta\Delta G_{PAK,O}$, however, is large (unfavorable) for the multiply substituted isomers (Figure 7e). Either the interaction of the ligands with the binding site must be weaker than their interactions with octanol or there must be a significant cost of reorganizing the binding site to accommodate the ligand.

The values of $\delta\Delta G_{W\rightarrow O}$ for the isosteric class II compounds vary much less than $\delta\Delta G_{W\rightarrow V}$. The range in $\delta\Delta G_{W\rightarrow O}$ is ~1.3 kcal/mol, corresponding to a factor of 10 in water–octanol partition coefficients. Both the $\delta\Delta G_{STAT}$ and $\delta\Delta G_{CAV}$ terms show only modest variation. In this case, the packing energy, $\delta\Delta G_{PAK,O}$, is negative for three of the four class II ligands (Figure 7e). Even though these values are negative relative to benzene, and not necessarily in an absolute sense, this result is nevertheless surprising. Apparently, these ligands interact more favorably with the protein interior than with octanol, relative to benzene and the other ligands. Two explanations for this may be proposed. First, these ligands all have very similar shapes, which are different from the other ligands used in the study. It is possible the binding site happens to complement this specific shape better than the time-averaged "binding site" offered by octanol. Presumably, this would require that all the isosteric ligands bind in identical or near-identical orientations, whereas some variability is observed (Morton & Matthews, 1995). Alternatively, the ligands may have more favorable interactions with the groups in the protein interior than with the octanol molecules. This could result from formation of specific electrostatic interactions between the heteroatoms of the class II molecules and functional groups such as the peptide bond. In the case of indole (but not the other class II ligands) the crystallographic analysis reveals formation of an NH···S hydrogen bond in the complex (Morton & Matthews, 1995).

Contribution of Solvent-Transfer Free Energies to Stability and Binding. A common method to interpret the contribution of hydrophobicity to the stability of nonpolar associations is to plot observed changes in stability or binding as a function of transfer free energy (e.g., Figure 5). Recently, Ben-Naim (1994) has argued that simple relationships between these quantities should not be expected and, when observed, are not meaningful except under special circumstances. We believe that transfer free energies do reflect an important *component* of stability, and our analysis reflects this belief. It also illustrates, however, that the binding strength of various ligands depends on several terms (cf. eq 8). Except in cases where all terms except the hydrophobic term drop out, plots of stability as a function of transfer free energy can be misleading. Figure 5c, for example, shows a simple linear relationship between stability and water–octanol transfer free energy, although the slope is only 0.5. Previously, similar results have led to the interpretation that a binding site can be more or less hydrophobic than octanol (Fersht, 1985). On the contrary, the analysis in Table 5B suggests this result is due to compensation of other terms in eq 8, such as $\delta\Delta G_{STAT}$ and $\delta\Delta G_{CAV}$. In the present case these terms and $\delta\Delta G_{W\rightarrow O}$ are all correlated with ligand size, and thus with each other as well, resulting in linear plots of slope other than unity in cases where any remaining terms

(in this case, $\delta\Delta G_{PAK,O}$) are near zero. Similar compensations are likely to occur in other cases where linear relationships are observed in plots of relative stability *versus* transfer free energy.

Whether the hydrophobic effect is or is not the dominant force in protein stability, it has long been the dominant idea in the protein folding literature. In using solvent-transfer free energies to model hydrophobic stabilization, the assumption is usually made that the interactions made by nonpolar groups within the protein are similar to those made in a nonpolar solvent [e.g., Baldwin (1986) and Lee (1993)]. Normally, this means that $\Delta G_{P:L} = \Delta G_{O:L}$ and $\Delta G_{CAV} = -\Delta G_{REORG}$ and that eq 8 becomes $\delta\Delta G_{OBS} = \delta\Delta G_{W\rightarrow O} + \delta\Delta G_{STAT}$. Because the present system contains a preformed cavity, ΔG_{REORG} is not offset by ΔG_{CAV}, and both must be considered explicitly. Doing so allows us to examine the assumption that $\Delta G_{P:L} = \Delta G_{O:L}$, assuming that ΔG_{REORG} is small. For one set of ligands, which have shapes similar to the preformed cavity, we find that the quantity $\delta\Delta G_{PAK,O} = \delta\Delta G_{P:L} - \delta\Delta G_{O:L} + \delta\Delta G_{REORG}$ is close to zero. This supports the idea that the interactions made by these ligands in the protein and in octanol are similar. Chemically similar ligands with different shapes, however, have values of $\delta\Delta G_{PAK,O}$ which are far from zero. In the accompanying paper, we structurally characterize these complexes to ask whether these values can be ascribed to the reorganization term or whether the assumption that $\delta\Delta G_{P:L} = \delta\Delta G_{O:L}$ breaks down. To anticipate the results, we conclude that the various reorganizations seen in the complexes are roughly isoenergetic and ascribe the differences in $\delta\Delta G_{PAK,O}$ to differences in $\delta\Delta G_{P:L}$ for the various ligands. In general, and particularly for other systems in which the binding site is not a preformed cavity, it becomes a semantic issue as to whether the origins of "packing" interactions lie in reorganization of the protein matrix or in different interactions in the final complex. The reason is that reorganization within the protein and the final interactions with the ligand are both dependent on the dynamic behavior of the protein and it is experimentally impossible to separate the two.

ACKNOWLEDGMENT

We thank many colleagues for discussions, particularly Drs. Doug Barrick, B. K. Lee, Hong Qian, John Schellman, and Brian Shoichet. One of the referees also provided some very useful comments, some of which are incorporated within the concluding section. We also thank Dr. F. W. Dahlquist and V. Feher for communication of results in advance of publication and Dr. John Keana for providing phenyl azide derivatives.

REFERENCES

Baldwin, E. P., Hajiseyedjavadi, O., Baase, W. A., & Matthews, B. W. (1993) *Science 262*, 1715−1718.
Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069−8072.
Ben-Naim, A. (1994) *Curr. Opin. Struct. Biol. 4*, 264−268.
Bigler, T. L., Lu, W., Park, S. J., Tashiro, M., Wieczorek, M., Wynn, R., & Laskowski, M., Jr. (1993) *Protein Sci. 2*, 786−799.
Bondi, A. (1964) *J. Phys. Chem. 68*, 441−451.
Burley, S. K., & Petsko, G. A. (1986) *J. Am. Chem. Soc. 108*, 7995−8001.
Connelly, P. R., Varadarajan, R., Sturtevant, J. M., & Richards, F. M. (1990) *Biochemistry 29*, 6108−6114.
Connolly, M. L. (1983) *J. Appl. Crystallogr. 16*, 548−558.
Creamer, T. P., & Rose, G. D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5937−5941.
Dill, K. A. (1990) *Biochemistry 29*, 7133−7155.
Eriksson, A. E., Baase, W. A., Wozniak, J. A., & Matthews, B. W. (1992a) *Nature 355*, 371−373.
Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P., & Matthews, B. W. (1992b) *Science 255*, 178−183.
Eriksson, A. E., Baase, W. A., & Matthews, B. W. (1993) *J. Mol. Biol. 229*, 747−769.
Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G., & Wells, J. A. (1986) *Science 233*, 659−663.
Fersht, A. (1985) *Enzyme Structure and Mechanism*, W. H. Freeman, San Francisco.
Gill, S. J., Nichols, N. F., & Wadso, I. (1975) *J. Chem. Thermodyn. 7*, 175−183.
Gill, S. J., Nichols, N. F., & Wadso, I. (1976) *J. Chem. Thermodyn. 8*, 445−452.
Hill, T. L. (1960) *Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading, MA.
Hine, J., & Mookerjee, P. K. (1975) *J. Org. Chem. 40*, 292−298.
Janin, J., & Finkelstein, A. V. (1990) *Protein Eng. 3*, 1−2.
Karpusas, M., Baase, W. A., Matsumura, M., & Matthews, B. W. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 8237−8241.
Kauzmann, W. (1959) *Adv. Protein Chem. 16*, 1−63.
Kellis, J. T., Jr., Nyberg, K., & Fersht, A. R. (1989) *Biochemistry 28*, 4914−4922.
Lee, B. (1991) *Biopolymers 31*, 993−1008.
Lee, B. (1993) *Protein Sci. 2*, 733−738.
Lim, W. A., & Sauer, R. T. (1989) *Nature 339*, 31−36.
Mark, H., & Tobolsky, A. V. (1950) *Physical Chemistry of High Polymeric Systems*, p 105, Interscience Publishing, Inc., New York.
Matsumura, M., & Matthews, B. W. (1989) *Science 243*, 792−794.
Matsumura, M., Becktel, W. J., & Matthews, B. W. (1988) *Nature* 334, 406−410.
Morton, A., & Matthews, B. W. (1995) *Biochemistry 34*, 8576−8588.
Pickett, S. D., & Sternberg, M. J. E. (1993) *J. Mol. Biol. 231*, 825−839.
Reiss, H. (1965) *Adv. Chem. Phys. 9*, 1−84.
Richards, F. M. (1974) *J. Mol. Biol. 82*, 1−14.
Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
Sandberg, W. S., & Terwilliger, T. C. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1706−1710.
Sangster, J. (1989) *J. Phys. Chem. Ref. Data 18*, 1111−1229.
Schellman, J. A. (1976) *Biopolymers 15*, 999−1000.
Shortle, D., Stites, W. E., & Meeker, A. K. (1990) *Biochemistry 29*, 8033−8041.
TRC Thermodynamic Tables, Hydrocarbons (1986) Thermodynamic Research Center, Texas A&M University, College State, TX.
Wiseman, T., Williston, S., Brandts, J. F., & Lin, L. N. (1989) *Anal. Biochem. 179*, 131−137.
Wolfenden, R., & Radzicka, A. (1994) *Science 265*, 936−937.
Yutani, K., Ogasahara, K., Tsujita, T., & Sugino, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4441−4444.
Zhang, X.-J., & Matthews, B. W. (1995) *J. Appl. Crystallogr.* (in press).

BI950297R