Journal of Solution Chemistry, Vol. 14, No. 2, 1985

Interactions between Terminally Substituted Amino Acids in an Aqueous and a Non-Aqueous Environment. Enthalpic Interaction Coefficients in Water and in N,N-Dimethylformamide at 25 °C

Hilary E. Kent,[1] Terence H. Lilley,[1,2] Peter J. Milburn,[1] Michael Bloemendal,[3] and Gus Somsen[2,3]

Received October 22, 1984; revised December 12, 1984

Enthalpies of dilution of the N-acetyl amides of glycine, L-alanine, L-valine, L-leucine, and L-phenylalanine, dissolved in N,N-dimethylformamide (DMF) as a solvent have been measured at 25 °C. The results obtained have been analyzed to give the enthalpic interaction (or virial) coefficients of the solutes and these are compared with information previously obtained in aqueous systems. There are marked differences in the interaction properties in the two solvents and, while the additivity approach of Savage and Wood is applicable to the solutes in water it is not suitable for representing the interactions in DMF. A correlation is presented between the enthalpic second virial coefficients in DMF and the propensity of side-chains to be in proximity in globular proteins.

KEY WORDS: Solute-solute interaction; enthalpies of dilution; substituted amino acids; enthalpic virial (interaction) coefficients; group additivity.

1. INTRODUCTION

The energetics involved in the weak (non-bonding) interactions of organic molecules in aqueous systems has attracted considerable attention for many years,[(1)] and more recently[(2)] some efforts have been directed towards the corresponding interactions in non-aqueous

[1]Department of Chemistry, The University, Sheffield, S3 7HF, U.K.

[2]To whom correspondence can be addressed.

[3]Department of Chemistry, Free University, De Boelelaan 1083, 1081 HV Amsterdam, Netherlands.

0095-9782/85/0200-0101$04.50/0

systems. In some contributions[3-11] from one of our laboratories, information has been presented on the behavior of some terminally substituted amino acids and peptides in water. The ultimate objectives of this latter work are related to problems in protein chemistry and in particular to

1. the tendency of some polypeptides to fold spontaneously into relatively well-defined structures[12], and

2. the propensity of peptide substrates to interact with the active sites of some enzymes.[13]

The first of these is a well-known problem and has been discussed at length in several places.[12,14-17] The second problem has not been considered as frequently and there is relatively little information on the energetics of such processes although this is an area which is attracting attention[18] from molecular graphic (*i.e.*, stereochemical) viewpoints.

In both of these areas it is clear that the problems involved are both complex and complicated and that the net energetics of the sociative[19] processes will have many component contributions. Some of these, at least, will be environmentally mediated and if, for example, one considers the first problem mentioned above, then we can state that in the early stages of protein biosynthesis, interactions between amino acid sub-units will be occurring in an essentially aqueous medium and consequently the nucleation stages of protein folding will be determined to some extent by this medium. However, as folding proceeds, water will be excluded from the interior of globular proteins and the sub-units in interior regions will be interacting in an environment which is largely non-aqueous. For any particular buried sub-unit in a given protein, the immediate environment will be complex but, from the wealth of information which is available on protein structures from X-ray structural determinations, broadly speaking, the interactions between sub-units will be occurring in molecular surroundings consisting of amidic and hydrophobic regions. The object of the present work was to obtain experimental results pertinent to the interactions occurring between some molecules with some of the features of polypeptide chains, in an amidic solvent. The molecules chosen were such that they contained groups representative of the hydrophobic class of amino acid side-chains since such groups are those which tend to prevail in the interior regions of globular proteins. The solvent chosen to represent the amidic-hydrophobic environment was dimethylformamide

(DMF). Our choice of this amide was partly pragmatic in that one of the collaborating groups has considerable experience[2,22-26] in its use, but it was felt it would be at least an appropriate starting point since it does contain some of the elements of protein interiors *viz.* polar and apolar groups.

2. EXPERIMENTAL

2.1. Preparation and Purification of Materials

The procedures used for the preparation of compounds are given below. For identification, the pmr chemical shifts δ were obtained at ambient temperature and are referred to either internal or external TMS, depending on the solvent used. The chemical shift data are presented in p.p.m., and in the parentheses following these we give (i) the integrated signal intensity, (ii) the nature of the signal [*s* denotes a singlet, *d* denotes a doublet, *q* denotes a quartet, *m* denotes a multiplet, *bs* denotes a broad singlet] and (iii) the assignment of the signal. In some instances, the coupling constant *J* is given. The abbreviations CMAW 120 and NEM refer to a chloroform:methanol:acetic acid:water mixture in the ratio (by volume) of 120:18:2:3, and N-ethylmorpholine, respectively.

2,1.1. N-Acetylglycinamide

Glycine ethyl ester hydrochloride was neutralized with triethylamine (1 equiv.) and acetylated with acetic anhydride (1 equiv.) in dry pyridine. The solvents were evaporated under reduced pressure and the product crystallized from ethylacetate/hexane followed by ether/hexane (yield 42%).

Anhydrous ethanol (200 ml) was saturated with dry ammonia with cooling. This was added to the above N-acetylglycine ethyl ester (30 g, 0.23 mol) and the mixture was then stirred for 48 h at room temperature. The residue was crystallized from ethanol/ether (13.8 g, 52%) *mp.* 138.5-139 ° (lit.,[27] 138-139.5 °); (Found C, 41.69; H, 7.10; N, 24.38; $C_4H_8O_2N_2$ requires: C, 41.40; H, 6.95; N, 24.13%), δ (D_2O) 3.91 (2H, *s*, αCH_2) 2.1 (3H, *s*, CH_3CO).

2,1.2. N-Acetyl-L-Alaninamide

N-Benzyloxycarbonyl-L-alanine was prepared as follows. L-Alan-

ine (60 g, 0.67 mol) was dissolved in 4*N* NaOH (160 ml) and cooled to 0 °C. Benzylchloroformate (110 ml, 1.1 equiv.) and 4N NaOH (160 ml) were added simultaneously over 1 h, stirred for 30 minutes, and warmed to room temperature. The aqueous liquors were awashed with ether (3×70 ml), cooled to 0 °C, and acidifed to pH 1 with conc. HCl. The oily solid was extracted into ether (3×150 ml) and the combined extracts washed with water (150 ml) and saturated brine (150 ml). The ethereal solution was dried over $MgSO_4$, filtered, evaporated to dryness, and the residue recrystallized from ethyl acetate by the addition of petroleum ether (102 g, 81%) *mp.* 83-84 °C (lit.,[28] *mp.* 86-87 °), δ ($CDCl_3$) 7.25 (5H, *m*, C_6H_5) 5.58 (2H, *d*, *J*=8Hz, $ArCH_2$) 5.78 (1H, *d*, *J*=8Hz, NH) 4.37 (1H, *q*, *J*=8Hz, αCH) 1.36 (3H, *d*, *J*=8Hz, βCH_3) 9.68 (1H, *s*, $CONH_2$). The primary amide was obtained using the following procedure.

N-Benzyloxycarbonyl-L-alanine (41.8 g, 0.2 mol) was dissolved in dry THF (200 ml) and cooled to -15 °C. To this was added NEM (22 ml, 0.2 mol) and isobutylchloroformate (26.4 ml, 0.2 mol) and the mixture was stirred at -15 °C for 5 minutes. Ammonia solution (0.88 S.G., 50 ml) was added slowly maintaining the temperature at -15 °C. The mixture was stirred for 30 minutes, then warmed to room temperature. The solvent was evaporated under reduced pressure and the residue dissolved in ethyl acetate, washed with 5% citric acid (100 ml), water (100 ml), 5% sodium bicarbonate solution (100 ml), saturated brine (100 ml), dried over $MgSO_4$ for 20 minutes, and filtered. Crystallization was induced by addition of petroleum ether (35 g, 83%) *mp.* 132-133 °C (lit.,[29] 133 °), δ (CD_3OD) 7.35 (5H, *m*, C_6H_5) 5.10 (2H, *s*, $ArCH_2$) 4.15 (1H, *q*, *J*=8Hz, αCH) 1.35 (3H, *d*,*J*=8Hz, βCH_3).

N-Benzyloxycarbonyl-L-alaninamide (12, 5 g, 0.055 mol) was dissolved in 80% glacial acetic acid and hydrogenated overnight in the presence of Pd/C catalyst (1 g). After uptake of hydrogen was complete the solution was filtered and cooled to 0 °C. To this was added dry pyridine (50 ml) and acetic anhydride (1.1 equiv.) and the solvents evaporated under reduced pressure. Ethyl acetate (100 ml) was added and evaporated under reduced pressure four times. The product was filtered from ethyl acetate and crystallized from ethanol/ether (6.45 g, 88%) *mp.* 162 ° (lit.,[30] 162 °C); R_F 0.26 (CMAW 120); (Found C, 46.3; H, 7.61; N, 21.53; $C_5H_{10}O_2N_2$ requires: C, 46.14; H, 7.76; N, 21.50%); $[\alpha]_D^{22}$ (*C*=1%, EtOH) -45 ° (lit.,[27] $[\alpha]_D^{22}$ 39 °. (*C*=1%, EtOH), δ (D_2O) 4.3 (1H, *q*, *J*=7Hz, αCH) 2.0 (3H, *s*, CH_3CO) 1.4 (3H, *d*, *J*=7Hz,βCH_3).

2,1.3. N-Acetyl-L-Leucinamide

The N-benzyloxycarbonyl derivative of L-leucine was prepared using the following route.

L-Leucine (131 g, 1 mol) was dissolved in 4*N* NaOH (250 ml) and cooled to 0 °C. Benzylchloroformate (175 ml, 1.1 equiv.) and 4N NaOH (250 ml) were added slowly over 45 minutes at 0 ° to pH 9-10. The system was stirred for 1 h and then warmed to room temperature. The solution was washed with ether (100 ml), acidified to pH 2 with 6*N* HCl (200 ml) and extracted into ethyl acetate. This was washed with saturated brine, dried over $MgSO_4$, filtered and evaporated under reduced pressure to a viscous oil (262.2 g, 99%), δ ($CDCl_3$) 7.24 (5H, *m*, C_6H_5) 5.18 (2H, *s*, $ArCH_2$) 7.18 (1H, *bs*, NH) 4.28 (1H, *m*, αCH) 1.6 (2H, *m*, βCH_2) 4.28 (1H, *m*, γCH) 0.98 (6H, *m*, γCH_3) 9.0 (1H, *bs*, CO_2H).

N-Benzyloxycarbonyl-L-leucine (53 g, 0.2 mol) was dissolved in THF (100 ml) and cooled to -15 °C. To this was added NEM (25.3 ml, 0.2 mol) and isobutylchloroformate (26.4 ml, 0.2 mol) and the whole was stirred at -15 ° for 5 minutes. Ammonia solution (0.88 S.G., 50 ml) was added at -15 °C. The mixture was stirred for 30 minutes at this temperature and then warmed to room temperature. The product was worked up as for the L-alanine derivative (40g, 80%) *mp.* 122-123 ° (lit.[31], 122-123 °). The required amide was obtained from this as follows.

N-Benzyloxycarbonyl-L-leucinamide (15.56 g, 0.081 mol) was dissolved in 80% glacial acetic acid (150 ml) in the presence of 5% Pd/C catalyst (0.5 g) and hydrogenated overnight. After uptake of hydrogen was complete the solution was filtered and cooled to 0 °C. To this was added pyridine (150 ml) and acetic anhydride (8.5 ml, 1.1 equiv.). The solution was evaporated under reduced pressure and the residue crystallized to constant melting point from ethanol (5.1 g, 36.5%) *mp.* 133-134 ° (lit.,[32] 133-4 °; R_F 0.45 (CMAW120); (Found C, 56.01; H, 9.17; N, 16.42; $C_8H_{16}O_2N_2$ requires: C, 55.79; H, 9.36; N, 16.27%), δ (D_2O) 4.3 (1H, *m*, αCH) 2.1 (3H, *s*, CH_3CO) 1.68 (1H, *m*, γCH) 1.60 (2H, *m*, βCH_2) 0.92 (6H, *m*, γCH_3).

The preparation of N-acetyl-L-phenylalaninamide has been described elsewhere.[7] The N-acetyl-L-valinamide had been prepared earlier.[3] N,N-dimethylformamide from Baker (Analyzed Reagent) was dried by storage over 4A molecular sieves (Baker) for at least seven days. GLC (column packed with 0.5% Na_3PO_4, 5% Polyclykol 1000, Merck, on Chromosorb GAW 80-100 mesh) indicated a purity better

than 99.5%. The amount of water was less than 0.015 mass %.

2.2. Calorimetric Procedures

Enthalpies of dilution were determined with a LKB 10700-2 batch microcalorimetric system. The output signal of the measuring cell was amplified and integrated using a Kipp BD 12 integrating recorder. Details of the experimental procedure have been described before.[23] In order to speed the measurements the method of subsequent dilutions[2,33] was used, in which, after the first dilution experiment, a maximal and known amount of solution in one of the compartments of the measuring cell is replaced by a known amount of pure DMF. Thus, in the second experimental solution is mixed with a highly diluted solution of the same kind. The procedure is repeated several times.

3. RESULTS

A compilation of the dilution experiments in DMF is given in Table I. The table presents the enthalpy change ΔH when n_A moles of solute at molality $m_{A,i}$ are mixed with n_B moles of solute at molality $m_{B,i}$ (or with pure DMF; $n_B = m_{B,i} = 0$) to give a solution with final molality m_f. The enthalpy change ΔH can be written as

$$\Delta H = n_A[H^E(m_f) - H^E(m_{A,i})] + n_B[H^E(m_f) - H^E(m_{B,i})] \quad (1)$$

when $H^E(m)$ denotes the excess enthalpy per mole of solute at molality m. The molar excess enthalpies of a solution of a single compound may be represented by[34]

$$H^E(m) = B_2^h m + B_3^h m^2 + \ldots \quad (2)$$

in which B_2^h, B_3^h, ... denote the pair, triplet, and higher enthalpic interaction coefficients of the solute.[4] Combination of Eqs. (1) and (2) gives

$$\Delta H/n_A = \sum_{n>1} B_n^h[(m_f^{n-1} - m_{A,i}^{n-1}) + n_A^{-1} n_B (m_f^{n-1} - m_{B,i}^{n-1})] \quad (3)$$

Enthalpic interaction coefficients B_2^h, B_3^h and, when possible, B_4^h

[4] B_2^h and B_3^h are equivalent to the symbols h_{AA} and h_{AAA} respectively, used by the Sheffield group.

Table I. Enthapies of Dilution in DMF at 25 °C [a]

$m_{A,i}$	n_A	$m_{B,i}$	n_B	m_f	ΔH	$\Delta\%$ [b]
			N-Acetylglycinamide			
0.3836	1.4629	8.4	8.4	0.3056	48.71	1.5%
0.4841	1.8055	10.5	10.5	0.3836	69.40	2.0%
0.3056	1.1668	4.6	6.8	0.2220	45.25	-0.3%
0.4522	1.5393	9.9	9.8	0.3521	60.27	0.0%
0.2611	1.0949	2.8	5.8	0.1748	48.22	3.7%
0.6159	2.2019	0.0	0.0	0.4841	98.20	3.5%
0.6159	1.5492	0.0	0.0	0.4522	84.40	-0.5%
0.2220	0.3267	1.3	4.8	0.0629	28.22	2.0%
0.3836	0.2895	0.0	0.0	0.0609	46.51	0.6%
0.4841	0.2864	0.0	0.0	0.0631	55.81	-1.2%
			N-Acetyl-L-alaninamide			
0.3628	1.5032	8.3	7.9	0.2966	46.39	3.7%
0.4114	1.7047	9.2	8.9	0.3348	49.82	0.0%
0.4482	1.8006	10.0	9.7	0.3628	53.69	-0.3%
0.2966	1.2327	4.5	6.5	0.2210	35.43	1.5%
0.5043	2.0366	11.6	10.8	0.4114	56.91	2.0%
0.6293	2.1936	0.0	0.0	0.5043	69.07	-0.2%
0.5676	2.0012	0.0	0.0	0.4482	69.97	-0.4%
0.2210	0.4566	0.0	0.0	0.0765	118.77	2.8%
0.2966	0.2111	0.0	0.0	0.0455	197.70	0.2%
0.3348	3.4296	1.8	7.2	0.0701	199.68	-1.1%
0.3628	0.1766	0.0	0.0	0.0402	246.96	-0.3%
			N-Acetyl-L-valinamide			
0.1211	0.4881	0.0	0.0	0.1086	7.70	-4.2%
0.0512	0.1997	0.5	1.2	0.0328	5.29	4.0%
0.0708	0.2992	1.0	1.6	0.0511	8.26	3.1%
0.0864	0.3652	1.3	1.9	0.0644	10.58	-1.4%
0.1211	0.5097	0.0	0.0	0.0951	16.60	-6.1%
0.0644	0.2753	0.6	1.5	0.0418	9.10	6.5%
0.0394	0.0807	0.0	0.0	0.0137	2.85	-2.1%
0.0787	0.3400	0.9	1.8	0.0538	12.12	5.3%
0.0418	0.0992	0.2	0.9	0.0152	3.71	0.4%
0.1172	0.4766	0.0	0.0	0.0864	20.04	2.2%
0.1031	0.4187	0.0	0.0	0.0708	18.78	2.9%
0.1086	0.4736	1.0	2.5	0.0706	23.86	-0.6%
0.0511	0.0860	0.3	1.1	0.0147	4.49	3.0%
0.1217	0.4745	0.0	0.0	0.0787	26.54	-1.0%
0.0706	0.1535	0.4	1.6	0.0243	9.35	-4.8%
0.1206	0.2896	0.0	0.0	0.0616	21.92	-4.0%
0.1031	0.2420	0.0	0.0	0.0369	20.81	-5.0%
0.1217	0.2580	0.0	0.0	0.0512	25.23	2.0%
0.1203	0.2315	0.0	0.0	0.0398	26.15	3.5%

Table I. Continued

$i_{A,i}$	n_A	$m_{B,i}$	n_B	m_f	ΔH	$\Delta\%$ [b]
			N-Acetyl-L-leucinamide			
0.0308	0.1255	0.3	0.7	0.0195	1.59	1.3%
0.0676	0.2867	0.6	1.5	0.0433	7.31	-0.3%
0.0990	0.4174	0.0	0.0	0.0676	13.36	-0.9%
0.0990	0.4194	0.9	2.2	0.0635	15.32	0.7%
0.0635	0.1517	0.3	1.4	0.0231	6.78	2.6%
0.1505	0.6821	1.4	3.4	0.0995	34.13	1.5%
0.3707	1.4863	7.9	7.8	0.2976	86.56	0.4%
0.1613	0.6019	1.5	3.6	0.0990	36.08	0.2%
0.2976	1.3118	4.4	6.5	0.2242	82.17	1.8%
0.2229	0.8888	0.0	0.0	0.1505	59.29	1.0%
0.0995	0.2305	0.5	2.2	0.0354	15.38	-0.3%
0.1332	0.1291	3.1	3.0	0.0681	8.63	-0.8%
0.4703	1.6596	0.0	0.0	0.3707	129.82	-1.0%
0.2840	0.1323	0.0	0.0	0.1332	17.64	-1.3%
0.4703	0.6104	0.0	0.0	0.2644	101.70	0.2%
			N-Acetyl-L-phenylalaninamide			
0.0548	0.2327	0.5	1.2	0.0354	4.21	-0.4%
0.0847	0.3558	0.8	1.9	0.0548	9.98	2.0%
0.1280	0.5369	1.3	2.9	0.0847	21.14	1.9%
0.1603	0.6654	1.6	3.6	0.1046	32.41	1.7%
0.3411	1.1831	7.8	7.8	0.2705	62.69	2.6%
0.2705	1.0843	4.3	5.9	0.2026	55.24	2.9%
0.2015	0.8120	0.0	0.0	0.1279	49.81	-2.9%
0.2398	0.9002	2.8	5.2	0.1603	58.94	-0.3%
0.4662	1.1971	0.0	0.0	0.3411	120.03	2.9%
0.4662	0.9222	0.0	0.0	0.2398	147.56	-1.3%

[a] Units: $m_{A,i}$ and m_f, mol-kg^{-1}; $m_{B,i}$, mmol-kg^{-1}; n_A, mmol; n_B, μmol; ΔH, mJ. [b] $\Delta\% = 100[\Delta h(\text{exp}) - \Delta H(\text{calc})] / \Delta h(\text{exp})$, where $\Delta H(\text{calc})$ is calculated from Eq. (3).

have been calculated by a least squares analysis of the results in Table I in terms of Eq. (3). Resulting values and their standard deviations are collected in Table II together with the enthalpic interaction coefficients of the compounds dissolved in water.[3,7] For the interaction coefficients in DMF (except B_3^h of N-acetyl-L-valinamide) the Student's *t*-test indicated a probability of at least 95% that their values are not zero. Due to the limited solubility of N-acetyl-L-valinamide in DMF, the measurements with this compound refer to molalities below 0.125 mol kg^{-1}. Consequently the standard deviations in its interaction coefficients are large. Also, the Student's t-test for B_3^h indicated a probabil-

Table II. Enthalpic Interaction Coefficients of Acetylaminoacid Amides[a]

Solute Aminoacid Group	B_2^h	B_3^h	B_4^h
		in DMF	
Glycine	-609(7)[b]	257(11)[b]	-
Alanine	-886(6)	293(12)	-
Valine	-1432(50)	484(314)	-
Leucine	-1149(11)	804(49)	-601(65)[b]
Phenylalanine	-982(20)	377(30)	-
		in Water	
Glycine	-220(3)	48(2)	-
Alanine	268(5)	22(4)	-
Valine	1259(15)	-	-
Leucine	1714(31)	434(57)	-180(31)
Phenylalanine	1049(18)	-	-

[a] Units: B_2^h, J-kg-mol^{-2}; B_3^h, J-kg^2-mol^{-3}; B_4^h, J-kg^3-mol^{-4}. [b] The number in parentheses is the standard deviation of the coefficient.

ity of only 80% that it is not zero. When the B_3^h and higher terms are ignored the value of B_2^h for N-acetyl-L-valinamide becomes -1358 J-kg-mol^{-2} with a standard deviation of 13 J-kg-mol^{-2}.

From Eq. (3) it follows that

$$\Delta_{dil}H(m_{A,i} \rightarrow m_f) = \Delta H / n_A - n_A^{-1} n_B \sum_{n>1} B_n^h (m_f^{n-1} - m_{B,i}^{n-1}) \quad (4)$$

where $\Delta_{dil}H(m_{A,i} \rightarrow m_f)$ is the molar enthalpy change on diluting a solution from initial molality $m_{A,i}$ to final molality m_f. Since

$$\Delta_{dil}H(m_i \rightarrow m_f) / (m_f - m_i) = [B_2^h + B_3^h(m_f + m_i) + B_4^h(m_f^2 + m_i^2 + m_f m_i) + \ldots \quad (5)$$

We give a graphical representation of the experimental results in DMF as a function of $(m_f + m_i)$ in Fig. 1 with enthalpies of dilution calculated according to Eq. (4). In this figure, the exceptional results for N-acetyl-L-valinamide emerge clearly. For the other compounds smooth curves are obtained.

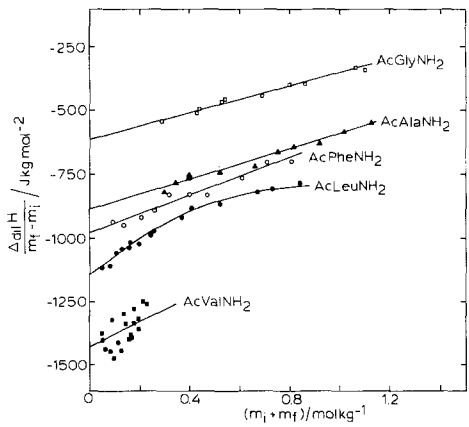

Fig. 1. $\Delta_{dil}H/(m_f - m_i)$ as a function of $m_f + m_i$ for the systems investigated in DMF.

4. DISCUSSION

Before the results obtained are discussed in detail, there are some qualitative points which can be made on the information presented in Table II. Firstly, if we consider the enthalpic pair interaction coefficients in water then it can be seen that the glycyl derivative exhibits a negative value but all of the other compounds have not only positive values but these increase in magnitude as the size and hydrophobicity of the amino acid side-chain increases. This trend is in marked contrast to that observed with DMF as solvent since in that case not only do all of the solutes have negative enthalpic pair interaction coefficients but the coefficients do not change in a monotonic way as the side-chain is extended and the most negative value is found for the valyl derivative. Secondly, it is apparent that the higher terms in the virial expansion are generally larger in magnitude for solutes dissolved in DMF than for those dissolved in water. Both features suggest that the interactions occurring between the solutes are both qualitatively and quantitatively different in the two solvents. It is also clear that, since a wide range of values are obtained for the interaction coefficients in both solvents, the intermolecular interactions of the solvated solutes are not dominated by hydrogen-bonding between amide groups on the solutes but also depend considerably on the nature of the amino acid side-chain. One other comment which should be made is that for the solutes in water, with the exception of the glycyl compound, since the pairwise enthalpic terms are positive, they give contributions to the net

interaction (as monitored by the free energetic terms[5]) which oppose association. The converse is the situation for all of the solutes in DMF since here all of the enthalpic terms are negative and are therefore thermochemically favorable.

Generally, as as implied in the Introduction, when one considers intermolecular interactions occurring between solutes, there are many features involved. One must, for example, recognize that one is concerned with solutes in a solvent and even when essentially 'infinitely dilute' solutions (*i.e.*, those in which the concentration of solute is sufficiently low so that inter-solute effects contribute negligibly) are considered, there must necessarily be some interplay between and modification of the properties of the solutes and the solvent peripheral to them. One must also recognize that for solutes such as those considered here, a range of conformations are possible, at least in principle, and each of these will have a characteristic although perhaps structurally ill-defined and certainly anisotropic solvation region. Consequently even when one is considering the properties of an 'isolated' solute in a solvent one must remember that a manifold of states is possible and that for fluxional molecules interconversion between them is relatively facile.

If we now turn to the situation where solutes are in propinquity, then it is apparent that a very large number indeed of different interactive situations are possible and these will all be solvent mediated to a greater or lesser extent. In other words, the experimental measures of solute-solute interactions will contain contributions not only from direct interactions between the solutes but also from, for example, conformational changes induced when solutes approach each other and from solvation shell perturbations occurring during the interactions. It would seem, therefore, that the problems associated with solute-solute interactions in solvents are multifarious and difficult to quantify. Consequently, at the present time, one must recourse to rather crude simplications of the molecular situations in the hope that some empirical rules may evolve which highlight the major features. One approximate procedure introduced by Savage and Wood,[35] which has been used to some effect is the 'additivity of groups' approach to solute-solute interactions. In this each solute is considered to be composed of a number of defined groups, and when two solutes interact it is assumed that the net interaction is the result of the (solvated) groups on one solute interacting with all of the (solvated) groups on the other solute.

This idea has been applied to aqueous solutions containing a

Table III. Enthalpic Pair Interaction Coefficient in DMF for Amides Containing Amino Acid Side-chains and the Normalized Interaction Frequency for Side-chain-Side-chain Interactions in Proteins

Side-chain	$-B_2^{h\,a}$	R
Alanine	886	0.0445
Valine	1432	0.1609
Leucine	1149	0.1472
Phenylalanine	982	0.1245

[a] Units: J-kg-mol^{-2}

range of solutes[3-11,35-42] and, although there are problems in some cases, it does seem to be at least semiquantitatively useful for solutes such as these considered here. In the Savage-Wood additivity of groups (SWAG) approach, no explicit solvent role is implicated although undoubtedly the intensive terms have major components from such sources implicated.

The SWAG approach has been used[23] to rationalize interactions occurring in DMF solutions and modification of it was necessary to implicitly include the solvent in the group formulation, even for relatively small solutes. However, it has also been recently shown[2] that when one considers amidic solutes with relatively long apolar side-chains then the interactions of these in DMF cannot be represented using the SWAG approach. Notwithstanding this, and in view of the utility of the approach for substituted amino acids in water as a solvent, we anticipated that it might also be applicable to small amidic solutes in DMF, and accordingly we pursued some analyses including the present and earlier data. Various attempts were made to represent the results using both the original and the modified SWAG treatments, but we found that our initial hopes were confounded and that the group additivity approach has no useful predictive ability for even relatively simple amidic solutes in DMF. This conclusion is disappointing but seems to be inescapable, and it appears that the treatment which works tolerably well for aqueous systems is inapplicable when DMF is the solvent. It was suggested in the earlier investigation of the interaction properties of long-chain amides in DMF[2] that the breakdown of the SWAG approach arose because of the dominance of one of the contributions to the net solute-solute interactions. If this departure stemmed from a fairly intensive side-chain-side-chain interaction, then

rather than the enthalpic coefficients having an approximately quadratic dependence on the side-chain length, a roughly linear dependence results. The experimental information seems to bear out this conclusion. It is apparent however that such a situation does not prevail with the presently studied compounds since, if one considers the non-aromatic compounds, only the trend in the enthalpic pairwise coefficients is not monotonic with increasing number of carbon atoms, but rather shows the lowest value for the valyl species. It would appear therefore that in the amidic solvent more structural discrimination is evident.

In view of this, we wondered if there was any correlation between the results obtained in DMF and the known tendency for certain amino acid residues to be found in proximity in globular proteins. This latter observation has been considered on several occasions and most recently by Roberts and Bohacek.[43] They surveyed the crystallographic coordinates of some 30 proteins and, from a statistical analysis, obtained a normalized measure of the frequencies of side-chain-side-chain interactions in these proteins. If one makes the presumably realistic assumptions that enhanced contacts between side-chains arise because of relatively favorable energetics and that the enthalpic measures obtained here also report on such interactions, then a correlation such as that referred to above should exist. The test of this is shown in Table III. It is apparent from this that for the small number of systems investigated there is some interrelation between the two measures of association, in that as the enthalpy of interaction between solute species becomes more negative, so too does the frequency of side-chain-side-chain contacts in proteins increase. It is especially striking that the most negative B_2^h value for the valine compound corresponds to the highest relative frequency for valine-valine contact.

This tentative conclusion encourages further experimentation, and the intention is that other peptidic solutes in DMF will be investigated by the Amsterdam group. The plan is that not only will homotactic (*i.e.*, like-solute-like-solute) interaction be studied, but also heterotactic (*i.e.*, like-unlike-solutes) interactions. While continuing their studies of aqueous systems, the Sheffield group has initiated two investigations, one into the properties of substituted amino acids and peptides in mixed water-amide solvent systems and another into the possible use of N-methylacetamide as a solvent to simulate protein interiors.

ACKNOWLEDGMENTS

We thank the following organizations for support in various ways: the Netherlands Foundation for Chemical Research (SON), the Netherlands Organization for the Advancement of Pure Research (ZWO), the Science and Engineering Research Council, the Agricultural and Food Research Council, and Roche Products Ltd. We also thank T. E. Leslie for discussions.

REFERENCES

1. J. J. Kozak, W. S. Knight, and W. Kauzmann, *J. Chem. Phys.* **48**, 675 (1968). This is a source of much of the early data.
2. M. Bloemendal and G. Somsen, *J. Solution Chem.* **13**, 281 (1984). This gives reference to earlier studies.
3. G. M. Blackburn, T. H. Lilley, and E. Walmsley, *J. Chem. Soc. Faraday Trans. I* **76**, 915 (1980).
4. G. M. Blackburn, T. H. Lilley, and E. Walmsley, *J. Chem. Soc. Chem. Comm.* 1091 (1980).
5. G. M. Blackburn, T. H. Lilley, and E. Walmsley, *J. Chem. Soc. Faraday Trans. I* **78**, 1421 (1982).
6. G. M. Blackburn, T. H. Lilley, and P. J. Milburn, *J. Chem. Soc. Faraday Trans. I*, in press.
7. G. M. Blackburn, H. E. Kent, and T. H. Lilley, *J. Chem. Soc. Faraday Trans. I*, in press.
8. K. Neiander, G. Olofsson, G. M. Blackburn, H. E. Kent, and T. H. Lilley, *Thermochimica Acta* **78**, 303 (1984).
9. G. M. Blackburn, T. H. Lilley, and P. J. Milburn, *J. Solution Chem.* **13**, 789 (1984).
10. T. E. Leslie and T. H. Lilley, *Biopolymers*, accepted for publication.
11. G. M. Blackburn, T. H. Lilley, and P. J. Milburn, *Thermochimica Acta*, accepted for publication.
12. See, *e.g.*, the articles in *Protein Folding*, R. Jaenicke, ed., (Elsevier, Amsterdam, 1980).
13. C. H. Hassall, W. H. Johnson, and N. A. Roberts, *Bioorganic Chem.* **8**, 299 (1979).
14. C. B. Anfinsen and H. A. Scheraga, *Advances in Protein Chem.* **29**, 205 (1975).
15. G. Nemethy and H. A. Scheraga, *Quart. Rev. Biophysics* **10**, 239 (1977).
16. C. Tanford, in *The Hydrophobic Effect*, (Wiley, New York, 1973).
17. J. T. Edsall and H. A. McKenzie, *Advances in Biophysics* **16**, 53 (1983).
18. See, *e.g.*, D. M. Blow, *Nature* **304**, 213 (1983) for some appropriate references.
19. E. A. Guggenhaim, *Trans. Faraday Soc.* **56**, 1159 (1960).
20. R. H. Wood, T. H. Lilley, and P. T. Thompson, *J. Chem. Soc. Faraday Trans. I* **74**, 1990 (1978).
21. T. H. Lilley, in *The Chemistry and Biochemistry of Amino Acids*, G. C. Barrett,

ed., (Chapman & Hall, London, 1984).
22. C. de Visser and G. Somsen, *J. Solution Chem.* **8**, 593 (1979).
23. M. Bloemendal and G. Somsen, *J. Solution Chem.* **12**, 83 (1983).
24. A. C. Rouw and G. Somsen, *J. Chem. Soc. Faraday Trans. I* **78**, 3397 (1982).
25. A. C. Rouw and G. Somsen, *J. Chem. Thermodynamics* **13**, 67 (1981).
26. A. C. Rouw and G. Somsen, *J. Solution Chem.* **10**, 533 (1981).
27. A. N. Kuntz and C. Niemann, *Biochim. Biophys. Acta* **50**, 520 (1961).
28. F. Machion, *Gazz. Chim. Ital.* **93**, 834 (1963).
29. H. Maluso, Y. Kawazoe, and M. Sato, *Bull. Chem. Pharm.* **18**, 1788 (1970).
30. H. Matsuo, Y. Kawazoe, M. Sato, M. Ohnishi, and T. Tatsuno, *Bull. Chem. Pharm.* **15**, 391 (1967).
31. O. K. Behrens and M. Bergmann, *J. Biol. Chem.* **129**, 578 (1938).
32. D. W. Ingles and J. R. Knowles, *Biochem. J.* **104**, 369 (1967).
33. P. T. Thompson, D. E. Smith, and R. H. Wood, *J. Chem. Eng. Data* **19**, 386 (1974).
34. H. L. Friedman, *J. Chem. Phys.* **32**, 1351 (1960).
35. J. J. Savage and R. H. Wood, *J. Solution Chem.* **5**, 733 (1976).
36. B. Y. Okamoto, R. H. Wood, and P. T. Thompson, *J. Chem. Soc. Faraday Trans. I* **74**, 1990 (1978).
37. T. H. Lilley and R. H. Wood, *J. Chem. Soc. Faraday Trans. I* **76**, 901 (1980).
38. A. L. Harris, P. T. Thompson, and R. H. Wood, *J. Solution Chem.* **9**, 305 (1980).
39. R. H. Wood and L. H. Hiltzik, *J. Solution Chem.* **9**, 45 (1980).
40. W. J. M. Heuvelsland, C. de Visser, and G. Somsen, *J. Chem. Soc. Faraday Trans. I*, **77**, 1191 (1981).
41. B. Y. Okamoto, R. H. Wood, J. E. Desnoyers, G. Perron, and L. Delorme, *J. Solution Chem.* **10**, 139 (1981).
42. I. R. Tasker and R. H. Wood, *J. Phys. Chem.* **86**, 4040 (1982); *J. Solution Chem.* **11**, 295 (1982); **11**, 469 (1982); **11**, 481 (1982); **11**, 729 (1982); **12**, 375 (1983).
43. D. C. Roberts and R. S. Bohacek, *Int. J. Peptide Protein Res.* **21**, 491 (1983). This gives reference to earlier work.