Bull. Inst. Chem. Res., Kyoto Univ. Vol. 66, No. 4, 1989

Characteristic Roles of Hydrated Water for 20 Amino Acid Residues in Stabilization of Globular Proteins

Motohisa OOBATAKE and Tatsuo OOI

Received September 20, 1988

Thermodynamic properties associated with hydrated water of proteins were computed on the native and extended three-dimensional structures and the average values of an accessible surface area (ASA), hydration free energy, enthalpy, and heat capacity of unfolding for every amino acid residue were obtained. Although ASA is approximately proportional to the size of a residue, the increment for the native conformation is different from one amino acid residue to another; that is, the residues are separated into two groups according to the degree of difference in ASA between native and extended conformations reflecting the location in the interior of the native protein. The first group includes small amino acid residues and polar and ionizable residues with about 50% decrease in ASA on folding. The second group contains hydrophobic residues and aromatic ones with about 80% decrease on folding. A difference in hydration free energy between both conformations also makes it possible to classify the residues into the two groups; a residue in the first group has a large negative value, while a residue in the second group has a small value close to 0. Each group consists of the same amino acid as ASA. Only differences are small amino acids (Gly, Ala, and Pro) in the second group and aromatic residues (Tyr and Trp) in the first group. The separation of the two groups is also observed by a difference of hydration heat capacity for both conformations. Amino acids in this group belong to the same group separated by ASA except for Cys which is assigned to the first group. Amino acid residues in the second group (Leu, Ile, Val, Met, Phe, and Cys) is more stabilized than the first group (Ser, Glu, Thr, and Asp) for mutant proteins of T4 lysozyme. This feature of stabilization is also observed for tryptophan synthase α subunit.

KEY WORDS: Hydration free energy/ Protein denaturation/ Accessible surface area/ Site-directed mutagenesis/

Although the hydration effects have been studied extensively for recognizing the important role in physical chemistry and biochemistry (1), it has not been easy to treat the hydration effect quantitatively due to the complicated water structure. Recently, we have developed a method to estimate thermodynamic properties of hydration, using an empirical linear relation between an accessible surface area and a property such as free energy, enthalpy or heat capacity for an atomic group. Proportional constants for 7 atomic groups occurring in proteins have been determined from thermodynamic experimental data on small molecules, assuming the additivity of contributions from constituent atomic groups (2). The temperature dependence of experimental unfolding free energy of several proteins could be accounted for by the calculation of free energy using the above method (3). Furthermore, association constants for the dimerization of proteins and proteolytic enzyme-inhibitor complexes have been estimated by removing the water molecules at the interacting regions of both molecules with loss in the hydration free energy and gain in the atomic

大島玄久, 大井龍夫 : Laboratory of Molecular Design of Physiological Functions, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611 Japan

interactions at contact sites in both molecules using the same method (4).

Characteristic thermodynamic properties of hydrated water for the 20 amino acid residues in globular proteins were computed utilizing the three-dimensional structures of proteins from the Protein Data Bank (5) and average values of hydration free energy, enthalpy, and heat capacity of unfolding for every amino acid residue were obtained (6). In the previous report (6), we focused the attention to the average values of unfolding quantities from a native structure to an extended structure. In this work, the average values of both structures will be shown, and then average hydration free energy will be examined in relation to the experimental free energy of unfolding of the mutant proteins.

METHOD

The hydration free energy of unfolding (ΔG_h^u) is defined as a difference in the hydration free energy between the native (ΔG_h^N) and random conformations (ΔG_h^R) at temperature T as follows;

$$\Delta G_h^u(T) = \Delta G_h^R(T) - \Delta G_h^N(T) . \qquad (1)$$

Similarly, the hydration enthalpy of unfolding (ΔH_h^u) is defined as a difference in the hydration enthalpy between the native (ΔH_h^N) and random conformations (ΔH_h^R) at temperature T as follows;

$$\Delta H_h^u(T) = \Delta H_h^R(T) - \Delta H_h^N(T) . \qquad (2)$$

Using atomic coordinates of a native conformation, hydration free energy and enthalpy at the standard temperature T_o (=25°C), can be computed from accessible surface areas of constituent atomic agoups (7 groups) as reported before (3); e.g.,

$$\Delta G_h(T_o) = \sum_{i=1}^{7} A_i g_i , \qquad (3)$$

where the accessible surface area (A_i) for i-th atomic group is computed using a water radius of 1.4 Å, and g_i is a parameter of free energy for the group. X-ray structures were available as native conformations. In order to calculate amounts of hydration, a random conformation represented by an extended structure of each protein was generated using the dihedral angles shown in Table I for every amino acid residue of a given amino acid sequence of the protein, and the structure had a low nonbonded energy according to the ECEPP algorithm (7). The N-and C-terminal residues were excluded from the statistics since the terminal residues are usually exposed to the solvent and behave differently from those in the middle of the sequence. The code names of 113 proteins from the Protein Data Bank (chain name) used for the statistics in the order of the residue number (Table I of ref. 6) are, 1PPT, 2RHV(4), 1CRN, 3RXN, 1FDX, 1OVO, 2OVO, 1TGS(Z), 4PTI, 2MT2, 2EBX, 1SN3, 1CTX, 2ABX, 3ICB, 1UBQ, 351C, 1CC5, 2B5C, 2GN5, 3FXC, 1PCY, 3CYT(O), 1RNT, 2FD1, 2CDV, 1REI(A), 2SSI, 1ACX, 1CPV, 1CCR, 1HMQ, 2RHE, 1CY3, 1BP2, 1P2P, 5RSA, 2CCY, 2AZA, 2LYZ, 1LZ1, 3FXN, 1HDS(A), 2HHB(A), 2MHB(A),

Table I. Dihedral angles of amino acid residues for random conformation taken from the minimum-energy conformations of the N-acetyl-N′-methylalaninamide (degree) (7).

	φ	ψ	χ^1	χ^2	χ^3	χ^4	χ^5	χ^6	χ^7
Ala	−155	160	59						
Asp	−155	160	58	− 83	−179				
Cys	−155	160	61	− 71					
Glu	−155	160	58	180	97	180			
Phe	−155	160	59	91					
Gly	−155	160							
His	−155	160	58	− 96					
Ile	−155	160	55	167	71	69			
Lys	−155	160	60	180	180	179	62		
Leu	−155	160	61	94	61	62			
Met	−155	160	58	180	180	60			
Asn	−155	160	61	− 47	−178				
Pro	− 75	160							
Gln	−155	160	57	180	−100	0			
Arg	−155	160	56	−179	−179	− 81	0	180	180
Ser	−155	160	60	− 60					
Thr	−155	160	−178	164	64				
Val	−155	160	63	64	54				
Trp	−155	160	57	74					
Tyr	−155	160	60	− 90	1				

The dihedral angles of the following sequences are changed to avoid the overlapping of atoms using ECEPP algorithm: ψ=140° of Thr and Val for Thr-Pro and Val-Pro, respectively, χ^2=−17° of Asn for Asn-X-Trp, χ^2=84° of Trp for Glu-X-Trp and Gln-X-Trp, χ^2=74° of Trp for Glu-Pro-Trp, and χ^2=64° of Trp for Leu-X-Trp and His-X-Trp.

1HDS(B), 1FDH(G), 2HHB(B), 2MHB(B), 1FX1, 2LHB, 2SOD, 1LH1, 2MBN, 1MBS, 3DFR, 2LZM, 3WGA(A), 1GCR, 1HMG(B), 2SGA, 2STV, 3SGB(E), 2ADK, 2ALP, 4SBV(A), 3FAB(L), 3GAP, 8PAP, 1FBJ(L), 1FB4(L), 1FBJ(H), 3FAB(H), 1MCP(L), 1MCP(H), 2PTN, 3RP2, 1FB4(H), 3PGM, 2PKA(A+B), 2RHV(3), 3CNA, 1IG2(H), 1EST, 2CGA, 1TIM, 2RHV(2), 2CAB, 1CAC, 2RHV(1), 1SBT, 1PYP, 1RHD, 1ABP, 5CPA, 3TLN, 2TBV(C), 2APP, 2APR, 1HMG(A), 1LDX, 4LDH, 4APE, 2GPD, 5LDH, 3GPD, 4ADH, 2CPP, 3PGK, 1CTS, 2GRS, 2TAA, and 8CAT. Following 12 proteins (1NXB, 1HIP, 1CYC, 2PAB, 1ECD, 2ATC(B), 1IG2(L), 2ACT, 2GCH, 4CHA, 2CYP, and 6API) were added for the data of the native confromations. In addition to the 12 proteins, 7 proteins (1INS(A+B), 2C2C, 1RNS, 1RN3, 1AZU, 2SNS, and 1MCG(1)) were available for the extended conformations.

RESULTS AND DISCUSSION

Average hydration quantities for the native and extended conformations computed at 25°C were as follows; accessible surface area (ASA^N, ASA^R), total free energy (ΔG_h^N, ΔG_h^R), total enthalpy (ΔH_h^N, ΔH_h^R), total heat capacity ($\Delta C_{p,h}^N$, $\Delta C_{p,h}^R$), free energy and enthalpy of the main chain (ΔGM_h^N, ΔGM_h^R, ΔHM_h^N, and ΔHM_h^R), and

Table II. Thermodynamic properties of hydration of amino acid residues averaged over 125 and 132 proteins for native and random conformations at 25°C, respectively. ASA^N, ASA^R: accessible surface area, ΔG_h^N, ΔG_h^R: hydration free energy, ΔH_h^N, ΔH_h^R: hydration enthalpy, $\Delta C_{p,h}^N$, $\Delta C_{h,p}^R$: hydration heat capacity, ΔS_h^N, ΔS_h^R: hydration entropy, ΔGM_h^N, ΔGM_h^R, ΔGS_h^N, ΔGS_h^R: hydration free energy for main chain and side chain, respectively, ΔHM_h^N, ΔHM_h^R, ΔHS_h^N, ΔHS_h^R: hydration enthalpy for main chain and side chain, respectively. Superscripts N and R mean the native and random conformations, respectively. The values in parentheses are standard deviations. Av means the average value over 20 amino acids. The units are: ASA (Å^2), energy, enthalpy (kcal mol^{-1}), heat capacity (kcal mol^{-1} K^{-1}), and entropy (cal mol^{-1} K^{-1}).

	ASA^N	ΔG_h^N	ΔH_h^N	$\Delta C_{p,h}^N$	ΔS_h^N
Ala	33.2 (31.8)	−0.06 (0.38)	− 0.96 (0.98)	8.45 (9.06)	− 3.02
Asp	62.4 (37.5)	−3.11 (2.46)	− 4.83 (3.50)	3.00 (5.56)	− 5.77
Cys	17.9 (24.2)	−0.27 (0.44)	− 0.55 (0.78)	1.31 (3.04)	− 0.94
Glu	81.0 (42.6)	−3.62 (2.38)	− 5.92 (3.56)	6.33 (6.94)	− 7.71
Phe	33.1 (40.8)	−0.28 (0.40)	− 1.17 (1.47)	8.36 (11.15)	− 2.99
Gly	29.2 (25.8)	−0.23 (0.44)	− 0.90 (0.90)	5.23 (5.55)	− 2.25
His	57.7 (44.2)	−2.18 (1.70)	− 4.21 (3.22)	10.94 (9.32)	− 6.81
Ile	28.3 (37.8)	0.07 (0.32)	− 0.78 (1.04)	8.58 (12.50)	− 2.85
Lys	107.5 (42.9)	−3.77 (1.89)	− 7.99 (3.28)	23.95 (11.23)	−14.15
Leu	31.1 (39.2)	0.07 (0.34)	− 0.85 (1.08)	9.24 (12.65)	− 3.09
Met	41.3 (49.5)	−0.10 (0.38)	− 1.17 (1.42)	9.78 (12.57)	− 3.59
Asn	60.5 (39.4)	−3.03 (2.37)	− 4.79 (3.49)	1.81 (4.91)	− 5.90
Pro	60.7 (39.8)	0.23 (0.36)	− 1.58 (1.04)	18.62 (13.11)	− 6.07
Gln	71.5 (43.9)	−3.15 (2.44)	− 5.19 (3.70)	3.84 (6.10)	− 6.84
Arg	94.5 (55.0)	−6.85 (4.07)	−11.23 (6.50)	11.57 (9.24)	−14.69
Ser	48.7 (34.5)	−2.36 (2.12)	− 4.16 (3.32)	8.68 (7.10)	− 6.04
Thr	52.0 (37.4)	−1.69 (1.77)	− 3.57 (2.94)	12.03 (9.82)	− 6.31
Val	28.1 (34.4)	0.04 (0.32)	− 0.77 (0.95)	8.14 (11.23)	− 2.72
Trp	39.5 (46.2)	−0.88 (1.09)	− 2.13 (2.47)	9.23 (11.91)	− 4.19
Tyr	50.4 (44.4)	−2.82 (2.63)	− 4.73 (4.21)	8.87 (9.49)	− 6.41
Av	51.4 (39.6)	−1.70 (1.41)	− 3.37 (2.49)	8.90 (9.12)	− 5.60

	ΔGM_h^N	ΔGS_h^N	ΔHM_h^N	ΔHS_h^N	$-T\,\Delta S_h^N$
Ala	−0.24 (0.41)	0.18 (0.19)	−0.39 (0.53)	− 0.57 (0.60)	0.90
Asp	−0.22 (0.34)	−2.88 (2.37)	−0.36 (0.41)	− 4.46 (3.37)	1.72
Cys	−0.17 (0.34)	−0.10 (0.25)	−0.22 (0.37)	− 0.32 (0.54)	0.28
Glu	−0.22 (0.34)	−3.40 (2.30)	−0.34 (0.39)	− 5.58 (3.44)	2.30
Phe	−0.13 (0.29)	−0.15 (0.22)	−0.19 (0.33)	− 0.98 (1.30)	0.89
Gly	−0.23 (0.44)	0.00 (0.00)	−0.90 (0.90)	0.00 (0.00)	0.67
His	−0.16 (0.31)	−2.02 (1.62)	−0.25 (0.36)	− 3.96 (3.09)	2.03
Ile	−0.13 (0.28)	0.20 (0.28)	−0.16 (0.31)	− 0.61 (0.87)	0.85
Lys	−0.25 (0.36)	−3.52 (1.83)	−0.36 (0.41)	− 7.63 (3.18)	4.22
Leu	−0.15 (0.31)	0.22 (0.29)	−0.19 (0.34)	− 0.66 (0.89)	0.92
Met	−0.13 (0.30)	0.03 (0.26)	−0.19 (0.34)	− 0.97 (1.24)	1.07
Asn	−0.22 (0.36)	−2.81 (2.26)	−0.33 (0.44)	− 4.45 (3.33)	1.76
Pro	−0.18 (0.29)	0.41 (0.29)	−0.30 (0.32)	− 1.27 (0.88)	1.81
Gln	−0.20 (0.34)	−2.95 (2.35)	−0.30 (0.38)	− 4.89 (3.57)	2.04
Arg	−0.20 (0.35)	−6.65 (3.99)	−0.29 (0.40)	−10.94 (6.37)	4.38
Ser	−0.28 (0.41)	−2.08 (1.95)	−0.45 (0.51)	− 3.71 (3.07)	1.80
Thr	−0.22 (0.38)	−1.47 (1.64)	−0.33 (0.43)	− 3.24 (2.76)	1.88
Val	−0.15 (0.29)	0.19 (0.25)	−0.19 (0.32)	− 0.58 (0.77)	0.81
Trp	−0.11 (0.28)	−0.77 (0.99)	−0.16 (0.29)	− 1.97 (2.34)	1.25
Tyr	−0.17 (0.33)	−2.66 (2.58)	−0.22 (0.37)	− 4.50 (4.14)	1.91
Av	−0.19 (0.34)	−1.59 (1.36)	−0.31 (0.41)	− 3.23 (2.41)	1.67

Table II. continued

	ASA^R	ΔG_h^R	ΔH_h^R	$\Delta C_{p,h}^R$	ΔS_h^R
Ala	104.0 (6.8)	− 0.58 (0.17)	− 3.20 (0.26)	22.71 (1.55)	− 8.79
Asp	132.2 (9.1)	− 6.10 (0.90)	− 9.39 (1.26)	5.64 (1.36)	−11.03
Gys	132.5 (7.6)	− 1.91 (0.21)	− 3.99 (0.30)	10.76 (1.20)	− 6.98
Glu	161.9 (9.0)	− 7.37 (0.77)	−11.61 (1.09)	9.47 (1.42)	−14.22
Phe	182.0 (12.0)	− 1.35 (0.21)	− 6.30 (0.48)	47.55 (3.19)	−16.60
Gly	73.4 (6.0)	− 0.82 (0.17)	− 2.37 (0.22)	10.14 (1.23)	− 5.20
His	165.8 (10.7)	− 5.57 (0.44)	−11.06 (0.77)	31.13 (2.32)	−18.41
Ile	171.5 (9.3)	0.40 (0.19)	− 4.63 (0.34)	50.71 (2.59)	−16.87
Lys	195.2 (9.0)	− 5.97 (0.20)	−13.04 (0.34)	41.68 (2.68)	−23.71
Leu	161.4 (10.3)	0.35 (0.18)	− 4.39 (0.34)	47.73 (3.45)	−15.90
Met	189.8 (9.1)	− 0.71 (0.17)	− 5.34 (0.32)	41.48 (2.34)	−15.53
Asn	134.9 (8.7)	− 6.63 (0.66)	−10.52 (0.97)	5.73 (1.19)	−13.05
Pro	135.1 (7.4)	0.56 (0.13)	− 3.54 (0.20)	42.60 (1.45)	−13.75
Gln	164.9 (9.5)	− 7.12 (0.59)	−11.46 (0.88)	7.44 (1.50)	−14.56
Arg	210.2 (10.7)	−12.78 (0.96)	−21.62 (1.43)	28.21 (1.99)	−29.65
Ser	111.4 (6.4)	− 6.18 (0.31)	−10.11 (0.45)	14.77 (1.32)	−13.18
Thr	130.4 (7.5)	− 3.66 (0.32)	− 7.97 (0.50)	28.29 (1.78)	−14.46
Val	143.9 (7.5)	0.18 (0.17)	− 3.93 (0.27)	40.78 (2.17)	−13.78
Trp	208.8 (12.2)	− 4.71 (0.31)	−11.19 (0.59)	47.26 (3.26)	−21.73
Tyr	196.4 (12.0)	− 8.45 (0.26)	−15.39 (0.53)	39.55 (3.07)	−23.28
Av	155.3 (9.0)	− 3.92 (0.37)	− 8.55 (0.58)	28.68 (2.05)	−15.53

	ΔGM_h^R	ΔGS_h^R	ΔHM_h^R	ΔHS_h^R	$-T\Delta S_h^R$
Ala	−1.15 (0.18)	0.56 (0.03)	−1.47 (0.21)	− 1.73 (0.08)	2.62
Asp	−0.77 (0.19)	− 5.33 (0.85)	−1.03 (0.22)	− 8.36 (1.19)	3.29
Cys	−0.74 (0.19)	− 1.17 (0.06)	−0.97 (0.22)	− 3.01 (0.14)	2.08
Glu	−0.80 (0.18)	− 6.57 (0.74)	−1.03 (0.21)	−10.58 (1.06)	4.24
Phe	−0.62 (0.18)	− 0.73 (0.07)	−0.88 (0.23)	− 5.42 (0.36)	4.95
Gly	−0.82 (0.17)	0.00 (0.00)	−2.37 (0.22)	0.00 (0.00)	1.55
His	−0.74 (0.19)	− 4.83 (0.38)	−0.98 (0.23)	−10.08 (0.69)	5.49
Ile	−0.79 (0.21)	1.19 (0.06)	−0.97 (0.23)	− 3.66 (0.18)	5.03
Lys	−0.80 (0.18)	− 5.17 (0.11)	−1.02 (0.21)	−12.02 (0.23)	7.07
Leu	−0.75 (0.18)	1.10 (0.07)	−0.99 (0.21)	− 3.40 (0.23)	4.74
Met	−0.77 (0.18)	0.06 (0.03)	−1.02 (0.21)	− 4.33 (0.20)	4.63
Asn	−0.66 (0.18)	− 5.97 (0.60)	−0.88 (0.21)	− 9.64 (0.90)	3.89
Pro	−0.36 (0.14)	0.92 (0.03)	−0.70 (0.13)	− 2.84 (0.09)	4.10
Gln	−0.78 (0.19)	− 6.34 (0.55)	−1.01 (0.22)	−10.46 (0.82)	4.34
Arg	−0.77 (0.18)	−12.02 (0.95)	−1.00 (0.22)	−20.62 (1.41)	8.84
Ser	−1.03 (0.18)	− 5.15 (0.24)	−1.32 (0.21)	− 8.79 (0.36)	3.93
Thr	−0.80 (0.18)	− 2.86 (0.22)	−1.02 (0.23)	− 6.94 (0.38)	4.31
Val	−0.79 (0.17)	0.97 (0.04)	−0.95 (0.20)	− 2.98 (0.14)	4.11
Trp	−0.69 (0.18)	− 4.02 (0.23)	−0.94 (0.20)	−10.24 (0.53)	6.48
Tyr	−0.61 (0.19)	− 7.84 (0.15)	−0.88 (0.23)	−14.51 (0.41)	6.94
Av	−0.76 (0.18)	− 3.33 (0.28)	−1.07 (0.21)	− 7.87 (0.49)	4.63

those of the side chain (ΔGS_h^N, ΔGS_h^R, ΔHS_h^N, and ΔHS_h^R), where superscripts, N and R represent the native and random conformations, respectively. The average values for both states of every amino acid residue together with the standard deviations are listed in Table II.

These values represent the characteristic feature of the hydration quantities of water soluble globular proteins at both states. ASA^R is proportional to the size of a residue for the extended conformations as illustrated in Fig. 1 with a slope of 18.6 Å^2/atom. There are, however, two groups having different slopes for the native

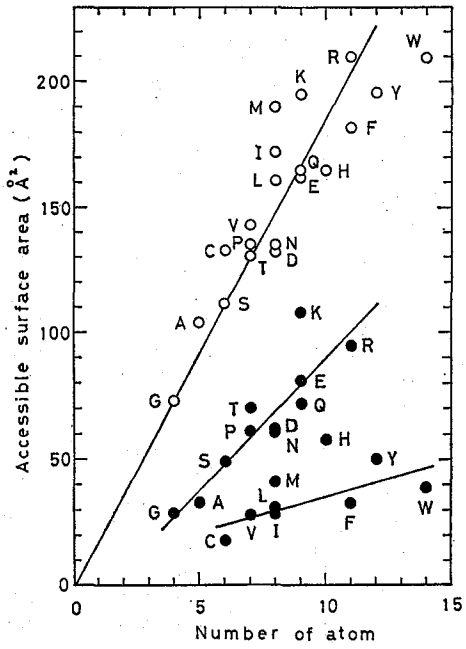

Fig. 1. Average accessible surface area of each amino acid residue is plotted against the number of heavy atoms. The values for the random conformation are indicated by open circles, and those for the native conformation by filled circles. The lines represent least-square analyses of the values.

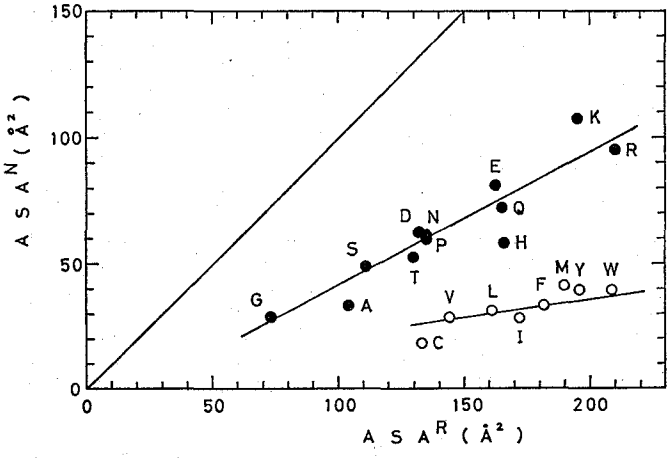

Fig. 2. Correlation between average accessible surface areas of each amino acid residue for native conformation, ASA^N, and those for random conformation, ASA^R. The lines represent least-square analysed analyses of the values. The diagonal line is shown as a reference. The residues which belong to the first group and the second group are shown by filled and open circles, respectively.

conformations; the slopes shown in Fig. 1 are 11 and 3 Å²/atom, respectively. Separation into the two groups is demonstrated more clearly by the plot of ASA^N for the native conformations against ASA^R for the random ones (Fig. 2). The first group

Table III. Thermodynamic properties of atoms of amino acid residues averaged over 125 (132) proteins for native (random) conformation at 25°C. ASA^N, ASA^R: accessible surface area for native and random conformations, respectively in Å^2, NA: the number of amino acid residues of all the proteins, Av: the average value over 20 amino acids except for CA atom in Gly, and CB atom in Ala. The order of atoms for tryptophan residues is shown and that for other residues is the order of the protein data bank. The values of polar atoms (N and O) are represented by bold letters.

ASA^N	NA	N	CA	C′	O	CB	CG	CD1	CD2	NE1	CE2	CE3	CZ2	CZ3	CH2
Ala	1681	**0.64**	4.25	0.09	**5.84**	22.08									
Asp	1184	**0.56**	4.17	0.10	**5.83**	15.43	1.40	**17.78**	**17.16**						
Cys	515	**0.37**	1.69	0.10	**4.48**	4.52	6.93								
Glu	997	**0.50**	3.68	0.08	**5.64**	10.16	16.36	1.92	**21.85**	**21.20**					
Phe	764	**0.33**	1.71	0.07	**3.52**	4.46	0.25	2.85	3.22	5.29	5.60	6.17			
Gly	1866	**1.27**	19.19	0.26	**8.71**										
His	508	**0.30**	2.86	0.09	**4.74**	7.86	0.74	**4.22**	9.78	17.00	**9.84**				
Ile	1011	**0.27**	1.16	0.04	**2.98**	1.32	3.68	8.69	10.52						
Lys	1293	**0.46**	3.68	0.08	**6.38**	8.49	10.75	16.94	29.96	**31.05**					
Leu	1653	**0.33**	1.24	0.05	**3.50**	3.26	1.51	10.27	11.31						
Met	333	**0.35**	1.88	0.07	**3.66**	4.35	5.39	9.01	17.06						
Asn	963	**0.56**	3.25	0.10	**5.87**	13.80	0.92	**15.50**	**21.01**						
Pro	992	**0.02**	4.78	0.07	**6.48**	17.92	21.55	10.75							
Gln	762	**0.38**	3.06	0.07	**5.52**	8.07	12.88	1.02	**18.76**	**21.93**					
Arg	703	**0.46**	2.76	0.08	**5.03**	7.05	8.20	14.22	**4.34**	4.74	**21.83**	**25.66**			
Ser	1740	**0.65**	5.15	0.10	**7.27**	22.19	**13.17**								
Thr	1427	**0.47**	3.35	0.08	**5.69**	6.54	**10.11**	26.03							
Val	1586	**0.29**	1.54	0.04	**3.61**	1.68	10.51	10.62							
Trp	322	**0.33**	1.43	0.06	**2.80**	4.54	0.29	5.37	0.54	**4.67**	1.04	2.33	6.98	4.07	6.25
Tyr	758	**0.34**	1.80	0.07	**4.15**	4.79	0.32	3.53	3.89	7.43	7.57	1.84	**14.45**		
Av	1052	**0.50**	3.03	0.09	**5.44**	8.07	6.06								

ASA^R	NA	N	CA	C′	O	CB	CG	CD1	CD2	NE1	CE2	CE3	CZ2	CZ3	CH2
Ala	1694	**2.58**	9.23	0.15	**24.75**	67.21									
Aps	1186	**0.90**	9.40	0.12	**20.50**	34.88	2.15	**33.70**	**30.52**						
Cys	517	**0.84**	8.96	0.15	**20.05**	32.36	70.04								
Glu	998	**0.90**	8.75	0.10	**20.84**	24.89	24.31	2.80	**40.77**	**38.46**					
Phe	770	**0.65**	9.51	0.14	**17.59**	31.13	1.49	10.79	10.11	31.02	33.79	35.78			
Gly	1887	**4.30**	42.14	0.92	**26.12**										
His	511	**1.05**	8.58	0.16	**19.37**	31.82	2.10	**8.75**	21.27	47.05	**25.58**				
Ile	1013	**0.85**	7.18	0.13	**20.71**	5.41	20.12	45.09	71.87						
Lys	1297	**0.90**	8.53	0.10	**20.67**	23.75	13.40	32.34	48.88	**46.69**					
Leu	1662	**1.00**	8.55	0.11	**19.40**	20.37	6.14	44.52	61.40						
Met	333	**0.86**	9.28	0.11	**20.30**	23.68	12.16	43.84	79.49						
Asn	973	**0.37**	9.13	0.12	**19.43**	33.76	1.38	**26.71**	**44.14**						
Pro	997	**0.03**	12.13	0.03	**12.30**	39.50	45.47	25.67							
Gln	764	**0.89**	8.73	0.13	**20.68**	23.85	23.69	1.71	**40.31**	**44.95**					
Arg	707	**0.85**	8.94	0.11	**20.29**	23.94	11.49	39.36	**13.41**	10.01	**26.68**	**55.19**			
Ser	1752	**1.96**	9.23	0.13	**23.59**	44.42	**32.12**								
Thr	1433	**1.05**	8.25	0.11	**20.33**	18.12	**20.53**	61.95							
Val	1595	**0.86**	6.59	0.10	**20.29**	7.70	57.09	51.31							
Trp	322	**0.93**	8.94	0.12	**18.24**	29.23	1.47	18.03	2.99	**24.61**	7.49	11.89	35.69	15.26	33.59
Tyr	761	**0.62**	9.79	0.15	**17.58**	31.28	1.53	10.54	10.87	32.47	29.31	9.37	**42.74**		
Av	1058	**1.35**	8.81	0.19	**20.77**	23.72	14.75								

contains small amino acid residues (Gly, Ala, and Pro), and polar and ionizable residues (Ser, Thr, Asp, Asn, Glu, Gln, Lys, Arg, and His) with about 50 % decrease in ASA on folding. On the other hand, the second group contains hydrophobic ones (Val, Leu, Ile, Met, and Cys) and aromatic groups (Phe, Tyr, and Trp) with about 80 % decrease in ASA on folding. That is, the amino acid residues in the second group are more buried than those in the first group. It is noteworthy mentioning that His shifts to the second group owing to its aromatic character and Ala also shifts to the second group. The residues in an increasing order of ASA^N for the native conformation are Cys, Val, Ile, Gly, Leu, Phe, and Ala. Although these residues belong to the second group (except for Gly and Ala) have extra side chain atoms compared to Gly and Ala, a larger residue does not always have a larger ASA in a native conformation. Contributions to ASA from every atomic group in an amino acid residue are shown in Table III; the end atomic group contributes most significantly, each group in the main chain has a similar value independent of the kind of amino acid, and Gly, Ala, and Pro have corresponding particular features. Generally, ASA of an atomic group for the native conformation is smaller than that for the extended conformation but two atomic groups, N in Asn and C′ in Pro are exceptions. This is originated from smaller values of ASA^R (0.37 for N in Asn and 0.03 for C′ in Pro) for the extended conformations in spite of the values of ASA^N (0.56 for N in Asn and 0.07 for C′ in Pro) which are close to average ones (0.50 for N and 0.09 for C′) for the native conformation.

The negative value of the average hydration free energy of every amino acid in Table II implies the gain in interaction energies with hydrated water; that is, an extended conformation is more stable than a native conformation in terms of hydration free energy. The plot of the average free energy for the native conformations against the extended ones (Fig. 3) demonstrates also the separation into the two groups; the first group having larger negative values and second group having smaller values close to 0. The residues included in each of the two groups are those in the corresponding group classified before; only differences are small amino acids (Gly, Ala, and Pro) in the second group and aromatic residues (Tyr and Trp) in the first group. The slope of the line for free energies at the native conformation is nearly the same as that of ASA for each of the two groups. In Fig. 3, the hydration free energy of unfolding is given by an amount from a point to the diagonal line. Although the first group has a value of -3.5 kcal mol^{-1} of hydration free energy of unfolding on average, the value of Arg (-5.96) is larger and that of Thr (-1.97) is smaller than the average. Four residues (Pro, Leu, Val, and Ile) above the diagonal line have the slightly positive value of ΔG_h^u, indicating that the unfolded state is favorable for most of the amino acid residues but unfavorable for these four residues due to hydration.

The feature of the average hydration enthalpy of every amino acid is similar to that of hydration free energy described before. All values of ΔH_h are negative whereas ΔG_hs of the four residues are positive. The plot of ΔH_h^N against ΔS_h^N for native conformations (Fig. 4) shows a linear dependence between the two quantities for each of the two groups, which contains the same amino acid as that in the

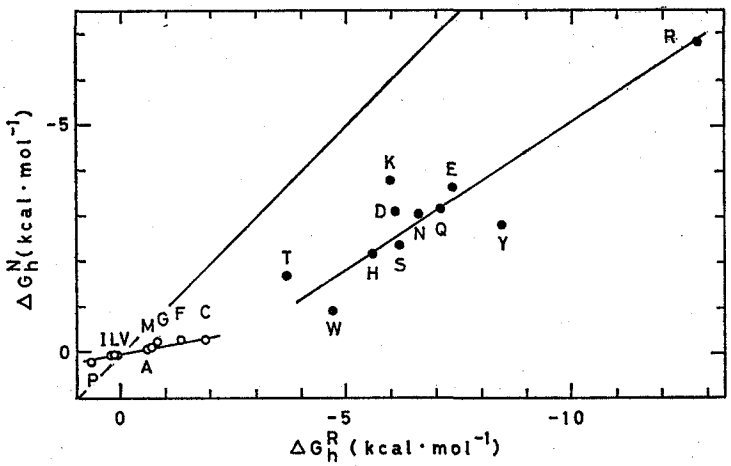

Fig. 3. Average hydration free energies of each amino acid residue for native conformation, ΔG_h^N, and those for random conformation, ΔG_h^R. The lines represent least-square analyses of the values. The diagonal line is shown as a reference. The residues which belong to the first group and the second group are shown by filled and open circles, respectively.

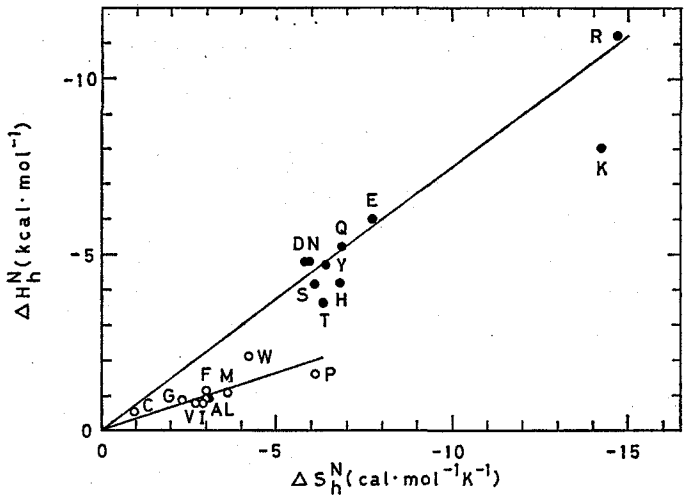

Fig. 4. Average hydration enthalpy of each amino acid residue, ΔH_h^N, is plotted against average hydration entropy, ΔS_h^N, for native conformation.

corresponding group for ΔG_h^N. The slopes $T_c=\Delta H_h^N/\Delta S_h^N$ are 747°K and 333°K for the first and second groups, respectively. This linear relation would be useful for the estimation of hydration entropy from hydration enthalpy.

The plot of $\Delta C_{p,h}^N$ vs. $\Delta C_{p,h}^R$ shown in Fig. 5 also demonstrates the separation of amino acid residues into the two groups. Each group contains the same amino acid as that of ASA except for Cys which is classified into the first group. The first group and the second group have nearly the same slopes 0.5 and 0.1 as those

for ASA, respectively. Residues in the second group which consists of hydrophobic and aromatic residues have larger values of unfolding heat capacity $\Delta C^u_{p,h}$ (30 to 40 kcal $mol^{-1}K^{-1}$) than those in the first group; however, Cys has a smaller values (9 kcal $mol^{-1}K^{-1}$) in spite of the large value of ΔASA^u.

The free energy of unfolding ΔG^u is a sum of the free energy of hydration ΔG^u_h and that of the chain ΔG^u_c. Recently, experiments of site-directed mutagenesis on Ile 3 of bacteriophage T4 lysozyme were reported (8). Ile 3 was substituted into 13 different amino acid residues and results on thermodynamic measurements and X-ray structure analysis on the mutants showed the hydrophobic stabilization at the

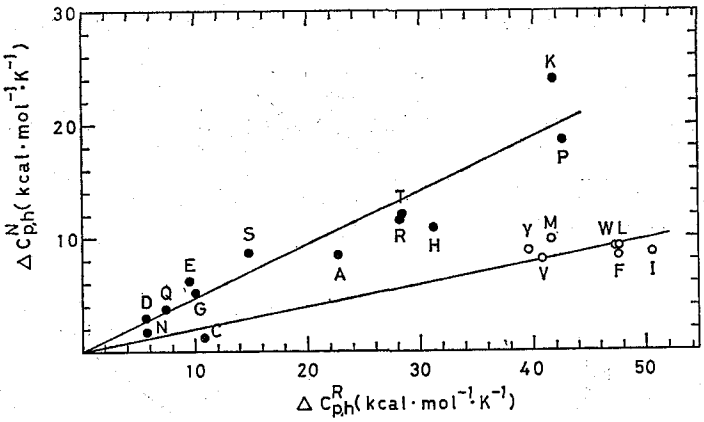

Fig. 5. Comparison of average hydration heat capacities of each amino acid residue for native conformation, $\Delta C^N_{p,h}$, and those for random conformation, $\Delta C^R_{p,h}$. The lines represent least-square analyses of the values.

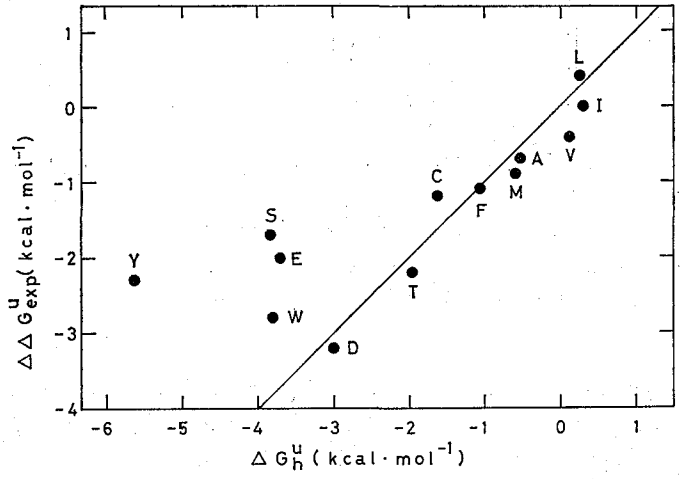

Fig. 6. Experimental free energy of stabilization ($\Delta\Delta G^u_{exp}$) of mutant T4 lysozymes (T_m= 64.7 °C at pH 6.5) is plotted against the hydration free energy of unfolding, ΔG^u_h, of individual amino acid residues. The free energies relative to wild-type (Ile 3) are plotted. The straight line represents both of the diagonal line and least-square analyses of the values.

Ile 3 position (8). In Fig. 6, ΔG_h^u is plotted against $\Delta\Delta G_{exp}^u$, which is the experimental difference in the free energy of unfolding between a mutant protein and wild-type one at the melting temperature of wild-type lysozyme (64.7°C at pH 6.5). Both quantities have a good correlation with each other and the slope of the straight line is about 1.0, indicating that ΔG_h^u contributes mainly to the overall stability of the protein. If we use water/ethanol transfer free energy instead of ΔG_h^u, the slope becomes 0.8 (8) indicating that the difference between the two slopes (0.2) originates from vacuum/ethanol transfer free energy since ΔG_h is regarded as vacuum/water transfer free energy. These results imply that the unfolding free energy of chain does not have a significant effect on the change in stabilization, while hydration free energy of unfolding is the major factor. The fact that Ile 3 in the native conformation ($ASA^N=19.9$ Å^2) is more buried than the average for Ile in globular proteins ($ASA^N=28.3$ Å^2) is related presumably to the above feature. Since hydrophobic residues of the second group have the average ASAs which are close to those of Ile 3, these residues may be easily replaced the position of Ile 3 with a small change in the free energy. On the contrary, a large deviation of Tyr from the diagonal line implies that the unfolding free energy of chain, 3 kcal mol^{-1}, must be involved in addition to the hydration free energy. Since Tyr has a much larger average value of ASA (50.4 Å^2) than that of Ile 3, Tyr cannot be replaced to the position of Ile 3 without a significant shift from the average. Indeed, it has been observed that the side chain of Tyr 3 protrudes into the solvent as revealed by the X-ray analysis (8). Another interesting point is that the second group (Leu, Ile, Val, Met, Phe, and Cys) and a small residue (Ala) stabilize the native conformation than the first group (Ser, Glu, Thr, and Asp) and aromatic residues (Tyr and Trp), and these groups are classified from the plot of hydration free energy (Fig. 3).

Other site-directed mutagenesis is mutation of tryptophan synthase α subunit; Glu 49 has been substituted into other 18 amino acid residues (9). In Fig. 7, free energy ΔG_{exp}^u at pH 7.0 determined experimentally is plotted against hydration free energy ΔG_h^u at 25°C; points distribute along a line having a slope of 2.0, indicating that ΔG_{exp}^u must consist of the hydration free energy ΔG_h^u and the chain free energy ΔG_c^u (10 to 15 kcal mol^{-1}). This considerable contribution of chain free energy ΔG_c^u is different from the case of T4 lysozyme. Interestingly, the second group (Ile, Leu, Met, Val, Phe, and Cys) stabilize the native conformation than the first group (Glu, Thr, Asp, Asn, Lys, and Ser) and the small residues (Ala, Pro, and Gly), and this classification is the same as that from the plot of ASA. Some aromatic residues (His and Trp), however, are located between the two groups. If unfolding free energy ΔG_{exp}^u is correlated with the optimal matching hydrophobicity (OMH) scale obtained from the frequency of amino acid replacements (10), aromatic residues (Tyr, Trp, and His) and ionizable ones (Asp and Glu) deviates from the straight line of the least-square fit (9). In this study, however, exceptional residues are small residues (Ala, Gly, and Pro) and aromatic ones (Tyr, Trp, and His); that is, the ionizable residues are on the line while the small residues are out of the line in contrast to the OMH scale. Experimentally, it is known that Glu 49 of tryptophan synthase is not ionized at pH 7.0, suggesting that the residue is located in the hydro-

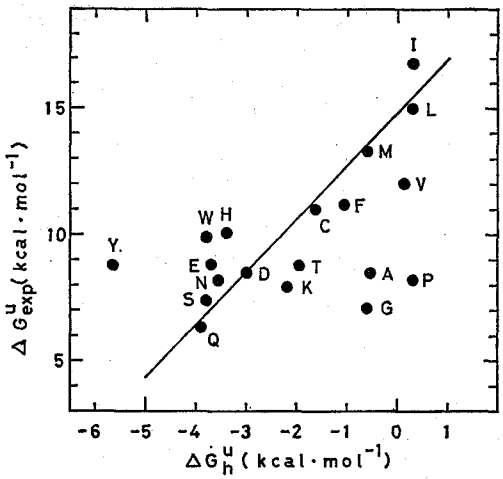

Fig. 7. Experimental unfolding free energy (ΔG^u_{exp}) in water of 19 proteins (T=25°C at pH 7.0) substituted by each of 19 amino acids at position Glu 49 of tryptophan synthase α subunit plotted against the hydration free energy of unfolding, ΔG^u_h, of individual amino acid residues. The straight line represents least-square analyses of the values.

phobic environment (9) as Ile 3 of T4 lysozyme. The more stabilization of the native structure by the second group than the first group for the two proteins would be attributed to the similar hydrophobic environment of Ile 3 and Glu 49 for T4 lysozyme and for tryptophan synthase, respectively. Therefore, the target residues (Ile 3 in T4 lysozyme and Glu 49 in tryptophan synthase) are easily replaced to other hydrophobic residues in the second group accompanying a small change in free energy. However, the aromatic residues and the small residues would induce a sizable conformational change, which gives rise to a significant change in free energy of unfolding unfavorable for the native conformation, i.e., destabilization is induced.

The results in this study represent quantitatively the characteristics of every amino acid in globular proteins, and the quantities obtained here may be useful for the construction of a predicted three-dimensional structure of a known amino acid sequence. For instance, the average value of every amino acid would be a useful measure for predicting the change in stability of a protein produced by the site-directed mutagenesis even when the three-dimentional structure of the protein is not known yet.

This work was supported by a Grant-In-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

(1) G. Ravishanker and D.L. Beveridge, "Theoretical Chemistry of Biological System", Elsevier, Amsterdam, 1986, pp. 423–494.

(2) T. Ooi, M. Oobatake, G. Némethy, and H.A. Scheraga, *Proc. Natl. Acad. Sci.* USA **84**, 3086–3090 (1987).

(3) T. Ooi and M. Oobatake, *J. Biochem.* (Tokyo) **103**, 114–120 (1988).
(4) T. Ooi and M. Oobatake, *J. Biochem.* (Tokyo) **104**, 440–444 (1988).
(5) F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyre, Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.* **112**, 535–542 (1977).
(6) M. Oobatake and T. Ooi, *J. Biochem.* (Tokyo) **104**, 433–439 (1988).
(7) G. Némethy, M.S. Pottle, and H.A. Scheraga, *J. Phys. Chem.* **87**, 1883–1887 (1983).
(8) M. Matsumura, W.J. Becktel, and B.W. Matthews, *Nature* **334**, 406–410 (1988).
(9) K. Yutani, K. Ogasahara, T. Tsujita, and Y. Sugino, *Proc. Natl. Acad. Sci.* USA **84**, 4441–4444 (1987).
(10) R.M. Sweet and D. Eisenberg, *J. Mol. Biol.* **171**, 479–488 (1983).