tions) are hypoplastic (*3*). Our observations suggest that these structural deficiencies may be the result of reduced cellular proliferation in the organogenesis phase, due to a direct action of ethanol (*14*). Clinical correlation of head size at birth with subsequent brain function has suggested that microcephaly is strongly related to mental retardation (*15*). Since we observed microcephalic growth in this study, the mental retardation seen in both fully and partially expressed FAS (*16*) may be the result of a direct inhibition by ethanol of neural growth early in gestation.

Our demonstration of ethanol-induced developmental retardation suggests that FAS may not be the result of maternally produced metabolites or altered maternal function. Whether the embryotoxic agent is ethanol itself or some other species produced by embryonic metabolism of ethanol is not yet clear. Current evidence, however, shows that embryos at this stage of gestation do not possess any ethanol-oxidizing or alcohol dehydrogenase activities (*17*). Although our results demonstrate that continuous exposure to high levels of ethanol exerts a direct toxic action on the developing embryo, the effects of short-term ethanol exposure have yet to be determined.

NIGEL A. BROWN*
Laboratory of Environmental Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, and, *Department of Pharmacology, George Washington University Medical Center, Washington, D.C. 20037*
EUGENIA H. GOULDING
Laboratory of Environmental Toxicology, National Institute of Environmental Health Sciences
SERGIO FABRO
Laboratory of Environmental Toxicology, National Institute of Environmental Health Sciences, and *Departments of Pharmacology* and *Obstetrics and Gynecology, George Washington University Medical Center*

References and Notes

1. Bureau of Alcohol, Tobacco and Firearms, *The Fetal Alcohol Syndrome: Public Awareness Campaign* (Government Printing Office, Washington, D.C., 1979).
2. P. Lemoine, H. Haronsseau, J.-P. Borteyru, J.-C. Menuet, *Quest Med.* **25**, 476 (1968); K. L. Jones, D. W. Smith, C. N. Ulleland, A. P. Streissguth, *Lancet* **1973-I**, 1267 (1973).
3. S. Clarren and D. W. Smith, *N. Engl. J. Med.* **298**, 1063 (1978).
4. Alcohol concentrations were chosen to mirror those blood alcohol concentrations observed following heavy alcohol consumption in humans [F. G. Hoffman, *A Handbook on Drug and Alcohol Abuse* (Oxford Univ. Press, New York, 1975), p. 102]. Since ethanol passes freely into the embryonic compartment [Y. A. Kesaniemi and H. W. Sippel, *Acta Pharmacol. Toxicol.* **37**, 43 (1975)], the embryo of a chronically alcoholic mother is likely to be exposed to such concentrations of alcohol for extended periods.
5. The day on which sperm was detected in vaginal smears was designated day 1. Embryonic ages were calculated under the assumption that fertilization occurred at the midpoint of the dark cycle (midnight).
6. D. A. T. New, P. T. Coppola, D. L. Cockroft, *J. Reprod. Fert.* **48**, 219 (1976). The medium was a homologous serum, immediately centrifuged and inactivated by heat.
7. D. L. Cockroft and P. T. Coppola, *Teratology* **16**, 141 (1977).
8. N. A. Brown, E. H. Goulding, S. Fabro, in preparation.
9. O. H. Lowry, N. J. Rosebrough, A. L. Farr, R. J. Randall, *J. Biol. Chem.* **193**, 265 (1951); U. Karsten and A. Wollenberger, *Anal. Biochem.* **77**, 464 (1977).
10. D. A. T. New, P. T. Coppola, D. L. Cockroft, *J. Embryol. Exp. Morphol.* **36**, 133 (1976).
11. Cell numbers were calculated by assuming 12 pg of DNA per embryonic cell [E. Kohler, H.-J. Merker, W. Ehmke, R. Wojnorowicz, *Naunyn-Schmiedeberg's Arch. Pharmacol.* **272**, 169 (1972)]. Approximate cell-cycle times were estimated from mean cell numbers of embryos taken over days $9^1/_2$ to $12^1/_2$ of gestation. Cell-cycle times increased from 6.2 hours at $9^1/_2$ days to 10.6 hours at 12 days, an increase compatible with the characterization of embryonic growth as Gompertzian [A. K. Laird, *Growth* **30**, 263 (1966)].
12. J. German, A. Kowal, K. Ehlers, *Teratology* **3**, 349 (1970); N. A. Brown, G. Shull, S. Fabro, *Toxicol. Appl. Pharmacol.*, in press.
13. N. A. Brown and S. Fabro, *Proceedings of the 26th Annual Meeting of the Society for Gynecological Investigation* **132**, A219 (1979).
14. Less common features of FAS are a range of minor and major malformations such as ventricular and atrial septal defects, cleft lip, cleft palate, and microphthalmia (*3*). These defects may also be the result of reduced growth of specific tissues during organogenesis, which is manifested later in gestation as dysmorphogenesis. Reduced cellular proliferation has been proposed as the mechanism of action of several teratogenic agents, particularly those which cause malformations such as cleft palate [W. J. Scott, in *Handbook of Teratology*, J. G. Wilson and F. C. Fraser, Eds. (Plenum, New York, 1977), vol. 2, pp. 81-98].
15. L. Crome, in *The Brain in Unclassified Mental Retardation*, J. B. Cavanaugh, Ed. (Churchill Livingstone, London, 1972), p. 284; A. Milunsky, *The Prevention of Genetic Disease and Mental Retardation* (Saunders, Philadelphia, 1975), pp. 19-50.
16. A. P. Streissguth, C. S. Herman, D. W. Smith, *J. Pediatr.* **92**, 363 (1978).
17. P. H. Pikkarainen and N. C. R. Raiha, *Pediatr. Res.* **1**, 165 (1967); N. C. R. Raiha, M. Koskinen, P. Pikkarainen, *Biochem. J.* **103**, 623 (1967); A. K. Rawat, *Ann. N.Y. Acad. Sci.* **273**, 175 (1976).
18. We thank the Audiovisual Services Department of George Washington University Medical Center for Fig. 1.
* Send reprint requests to N.A.B. at George Washington University Medical Center.

9 March 1979; revised 11 June 1979

Water, Protein Folding, and the Genetic Code

Abstract. *The absolute affinities of amino acid side chains for solvent water closely match their relative distributions between the surface and the interior of native proteins and are associated with a remarkable bias in the genetic code.*

Many processes of biological "recognition" require the stripping away (at least in part) of solvent water from interacting groups. The mutual affinities therefore reflect in part the ease with which they can be removed from solvent water, in addition to any specific forces of attraction or repulsion that may be present. We now report the free-energy changes associated with the removal of side chains of common amino acids from solvent water. These changes resemble the relative distributions of the amino acids between the surfaces and the interiors of native globular proteins, and are associated with a sharp bias in the genetic code.

The affinity of a compound for watery surroundings can be expressed quantitatively in terms of its free energy of transfer from the dilute vapor phase, in which intermolecular forces are virtually absent, to an aqueous solution so dilute that solute-solute interactions can be neglected. Results obtained for many compounds suggest that this measure of the

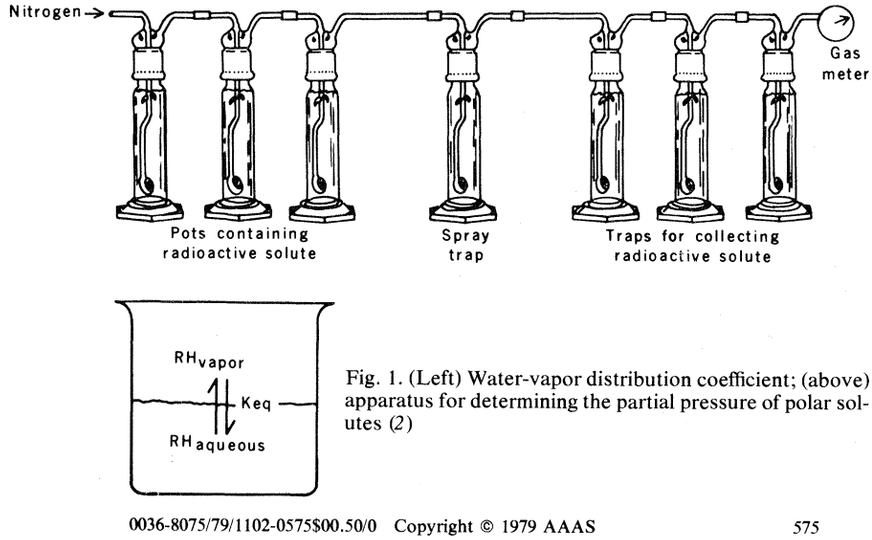

Fig. 1. (Left) Water-vapor distribution coefficient; (above) apparatus for determining the partial pressure of polar solutes (*2*)

0036-8075/79/1102-0575$00.50/0

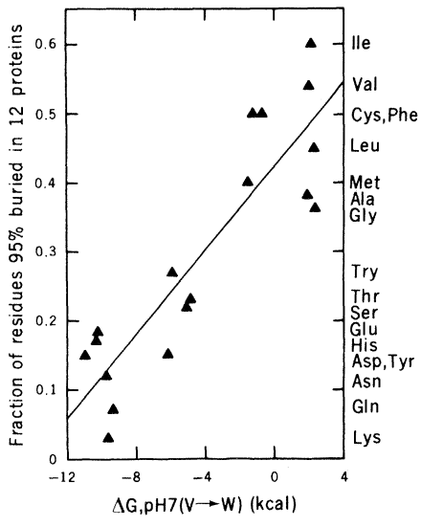

Fig. 2. Relative tendencies of amino acid residues to be inaccessible to solvent, observed in 12 crystalline proteins (*8*), as a function of the hydration potentials of the corresponding side chains. On the ordinate, the number of residues of each amino acid that are inaccessible to solvent over 95 percent of their surface areas is expressed as a fraction of the total number of residues of that amino acid that are present in all 12 proteins taken together.

hydrophilic character of complex molecules can be treated as an approximately additive function of their constituent groups (*1*). The peptide bond is unusual in the degree to which it is stabilized thermodynamically by interactions with solvent water (*2*). It would be useful to have similar information about the side chains of amino acids that are commonly found in proteins.

It is not possible at present to determine the partial pressures of aqueous amino acids, because of their extremely low volatility. Nor did this seem desirable since free amino acids, unlike internal amino acid residues in polypeptides, are zwitterionic in aqueous solution. Free energies of solvation of charged ammonium and carboxylate groups in water are very large, in the neighborhood of −70 to −80 kcal/mole for each group (*3*) as compared with about −10 kcal/mole for a simple peptide group (*2*). The disruptive local influence of a pair of charged groups might be expected to modify the apparent solvation properties of different side chains to an indeterminate extent, especially in view of the anomalous effects of chain length on the volume changes that accompany dissociation of normal carboxylic acids in water (*4*). To obtain results that might be more closely comparable with the behavior expected of amino acid residues in proteins, we examined the behavior of amino acid side chains alone, without substituents of any kind.

In most cases, a dynamic technique was used to measure equilibria of distribution of side chains, with radioactivity to detect the solute in the vapor phase (Fig. 1) (*2*); the distribution of 3-methylindole was determined spectrophotometrically, with cuvettes of 10-cm light path to detect the solute in the vapor phase (*5*). Measurements were made at 25°C, in the presence of a buffer appropriate to maintain the solute in an uncharged state, and sufficient KCl to adjust the ionic strength to 0.1. Other values, obtained under comparable conditions by other investigators, are shown in Table 1.

Table 1. Hydration potentials of amino acid side chains at 25°C. See Table 2 legend for abbreviations of amino acids.

Side chain	Amino acid residue	$K_{eq} = \frac{(RH_{H_2O})^*}{(RH_{vap})}$	α†	Hydration potential‡ (kcal/mole)
Hydrogen	Gly	1.7×10^{-2}	1	+2.39
Isobutane	Leu	2.1×10^{-2}	1	+2.28
n-Butane	Ile	2.6×10^{-2}	1	+2.15
Propane	Val	3.5×10^{-2}	1	+1.99
Methane	Ala	3.7×10^{-2}	1	+1.94
Toluene	Phe	3.6	1	−0.76
Methanethiol	Cys	8.1	1	−1.24
Ethyl methyl sulfide	Met	1.2×10^{1}	1	−1.48
Ethanol	Thr	3.9×10^{3}	1	−4.88
Methanol	Ser	5.2×10^{3}	1	−5.06
3-Methylindole	Trp	2.1×10^{4}	1	−5.89
4-Cresol	Tyr	3.1×10^{4}	1	−6.11
Propionamide	Gln	7.9×10^{6}	1	−9.38
n-Butylamine	Lys	1.6×10^{3}	1.6×10^{-4}	−9.52
Acetamide	Asn	1.3×10^{7}	1	−9.68
Propionic acid	Glu	5.5×10^{4}	1.8×10^{-3}	−10.19
4-Methylimidazole	His	3.2×10^{7}	9.1×10^{-1}	−10.23
Acetic acid	Asp	8.1×10^{4}	7.6×10^{-4}	−10.92

*RH represents the solute (acetamide in the case of Asn, for example) whose distribution coefficient was measured. The values for 3-methylindole, 4-methylimidazole, acetic acid, acetamide, and propionamide were determined by us. Other values were taken from a compilation of results (*2*), the value for ethyl methyl sulfide being interpolated from closely similar values for diethyl sulfide and dimethyl sulfide. Equilibria are for uncharged solutes. †α is the fraction of solute that is not ionized in aqueous solution, buffered at *p*H 7.0. Values for pK_a were assumed to be equivalent to the apparent pK_a values of the side chains, observed with the free amino acids (*13*), that is, lysine, 10.79; histidine, 6.00; glutamic acid, 4.24; and aspartic acid, 3.86. The pK_a values for the isolated side chains gave somewhat different values for α, but the resulting effects on the slope and correlation coefficient of Fig. 2 were insignificant. ‡Hydration potential is $-RT\ln(K_{eq}/\alpha)$.

Table 2. Code letters in messenger RNA (mRNA) and DNA. Abbreviations for amino acid residues are: Gly, glycine; Leu, leucine; Ile, isoleucine; Val, valine; Ala, alanine; Phe, phenylalanine; Cys, cysteine; Met, methionine; Thr, threonine; Ser, serine; Trp, tryptophan; Tyr, tyrosine; Gln, glutamine; Lys, lysine; Asn, asparagine; Glu, glutamic acid; His, histidine; and Asp, aspartic acid. Abbreviations for code letters are: G, guanine; U, uracil; C, cytosine; A, adenine; and T, thymine.

	Residue	Second code letter in	
		mRNA	DNA
	Gly	G	C
	Leu	U	A
	Ile	U	A
	Val	U	A
	Ala	C	G
	Phe	U	A
	Cys	G	C
	Met	U	A
	Thr	C	G
	Ser	C(G)	G(C)
	Trp	G	C
	Tyr	A	T
	Gln	A	T
	Lys	A	T
	Asn	A	T
	Glu	A	T
	His	A	T
Most hydrophilic	Asp	A	T

In order to generate a scale of free energies of transfer from the vapor phase to neutral aqueous solution, it seemed appropriate to correct for the influence of ionization of acidic and basic side chains under physiological conditions. Ionized forms of the side chains can be considered totally nonvolatile for present purposes, in view of the large negative free energies of hydration of ammonium and carboxylate groups. Equilibria observed for the transfer of uncharged side chains, from the vapor phase to water, were therefore divided by the fraction of each acidic or basic side chain present in uncharged form at *p*H 7. The resulting free energies (Table 1), designated for brevity "hydration potentials," span a range of about 13 kcal for the 18 amino acid side chains that we studied.

Protein structures have now been solved in sufficient numbers so that it is possible to use our findings to test the strength of subjective generalizations (*6*) that the polar character of amino acid

side chains determines their tendencies to be found at the surface of globular proteins. A rolling sphere model, proposed by Lee and Richards (*7*), provides a reasonable semiquantitative basis for determining the exposure to solvent of the individual atoms of a protein whose configuration has been determined by exact structural methods. Applying this method to 12 crystalline proteins, Chothia (*8*) has calculated the number of residues of each amino acid that are inaccessible to solvent over 95 percent of their individual surface areas, expressed as a fraction of the total number of residues of that amino acid, which are present in all 12 proteins taken together. In Fig. 2, this fraction is plotted as a function of the hydration potentials in Table 1. Amino acid residues falling on the lower half of one scale, also fall on the lower half of the other. There are no exceptions (*9*).

This close relationship led us to question whether there might be any correspondence between the hydration potential of the various side chains and the nucleic acids that serve as their genetic determinants. The amino acid code (*10*) has been found to be moderately degenerate at the first position of the codon and highly degenerate at the third position of the codon, in the sense that several alternative bases serve to code for the same amino acid at these positions; but each amino acid is uniquely associated with the presence of a single base at the second position of the codon except serine (two alternative bases). In our study, the 18 side chains, ranked in order of increasing hydration potential, can be divided into two equal groups (Table 2). It is evident from inspection that the observed distribution of code letters is nonrandom. Thymine, for example, serves as the second code letter for seven amino acids. Every one of them has a hydrophilic side chain, and all are clustered near the origin of Fig. 2. Estimated conservatively (*11*), the probability that this might occur by chance is in the neighborhood of .0045.

The present scale of hydration potentials illustrates the probable importance of amino acid side chain interactions with solvent water as a factor in determining the overall configurations of proteins. It is not apparent that any simple physicochemical basis exists for the observed correlations with the modern genetic code, but it seems natural to suppose that coding similarities between amino acids with similar physical properties may have tended to offset the disruptive effects of mutation on the structural stability of proteins during their evolution (*12*). Figure 2 provides quantitative support for the widely held view that mutations that would result in the introduction of hydrophilic amino acids, at interior locations that were previously hydrophobic, are expected to be especially damaging. The observed distribution of code letters appears to minimize the likelihood of these events.

R. V. WOLFENDEN
P. M. CULLIS
C. C. F. SOUTHGATE
Department of Biochemistry, University of North Carolina, Chapel Hill 27514

References and Notes

1. J. A. V. Butler, *Trans. Faraday Soc.* **33**, 229 (1937); J. Hine and P. K. Mookerjee, *J. Org. Chem.* **40**, 292 (1975).
2. R. Wolfenden, *Biochemistry* **17**, 201 (1978).
3. P. Kebarle, in *Environmental Effects on Molecular Structure and Properties*, B. Pullman, Ed. (Reidel, Dordrecht, 1976), pp. 81–94.
4. W. Kauzmann, A. Bodansky, J. Rasper, *J. Am. Chem. Soc.* **84**, 1777 (1962).
5. The absorption spectrum observed for 3-methylindole in the vapor phase was almost identical with its spectrum in cyclohexane. Its concentration in the vapor phase was therefore estimated from its extinction coefficient (ϵ) determined in cyclohexane in separate experiments: $\epsilon_{267} = 6.74 \times 10^3$.
6. W. Kauzmann, *Adv. Protein Chem.* **16**, 1 (1959); M. Perutz, *J. Mol. Biol.* **13**, 646 (1965).
7. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
8. C. Chothia, *ibid.* **105**, 1 (1976).
9. The coefficient for linear correlation, for the 18 amino acids included in Fig. 2, is 0.90 ($P \leq .00001$). An exact value for the hydration potential of the side chain of Arg remains to be determined. Preliminary results indicate that it is even more negative than that of Asp, consistent with its extreme tendency to be exposed at the surface of globular proteins in which it is found (*8*). Earlier scales of hydrophobicity that were based on free energies of transfer of amino acids near their isoelectric points to ethanol or dioxan [Y. Nozaki and C. Tanford, *J. Biol. Chem.* **246**, 2211 (1971)], or to the surface of their aqueous solutions [H. B. Bull and K. Breese, *Arch. Biochem. Biophys.* **161**, 665 (1974)], when used as abscissas in plots similar to that of Fig. 2, exhibit correlation coefficients of 0.20 and 0.58, respectively. Some of the differences between these various scales are almost certainly due to the special solvation properties of zwitterionic amino acids.
10. M. W. Nirenberg, O. W. Jones, P. Leder, F. C. Clark, W. S. Sly, S. Petska, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 549 (1963). The incidence of similar code words for similar amino acids, noted in a qualitative sense by these authors, has been a subject of commentary and speculation [see, for exam → F. H. C. Crick, *J. Mol. Biol.* **38**, 367 (1968); C. Woese, *ibid.* **43**, 235 (1969)].
11. Comparing the dichotomy hydrophilic/not-hydrophilic with the dichotomy T/not-T, the significance level by Fisher's exact test is .0011. One such test was performed with each of the four bases, so that we estimate that the real significance is about four times larger. Regarding the overall pattern of the distribution in Table 2, including all four bases, the nonrandomness of the distribution was confirmed by the Kruskal-Wallis test: the P value was approximately .005. We thank Drs. R. C. Grimson and R. C. Elston, Department of Biostatistics, University of North Carolina, for advice in this matter.
12. R. M. Sonneborn, in *Evolving Genes and Proteins*, V. Bryson and H. J. Vogel, Eds. (Academic Press, New York, 1965), pp. 377–397.
13. J. T. Edsall and J. Wyman, *Biophysical Chemistry* (Academic Press, New York, 1958).
14. Supported by NIH grant GM-18325 and NSF grant PCM-7823016.

5 June 1979; revised 15 August 1979

Candidiasis: Detection by Gas-Liquid Chromatography of D-Arabinitol, a Fungal Metabolite, in Human Serum

Abstract. *D-Arabinitol was identified as a major metabolite of* Candida *species in human subjects. Gas-liquid chromatography was used to measure the concentration of D-arabinitol in serum. The study included subjects who were healthy and cancer patients who had proven invasive candidiasis or were colonized with* Candida. *D-Arabinitol concentrations greater than 1.0 microgram per milliliter were found in serum from patients with invasive infection. This technique may prove valuable in the diagnosis of invasive candidiasis.*

Invasive candidiasis is a common, life-threatening infection in immunosuppressed hosts, such as transplant recipients or cancer patients (*1*); however, it is difficult to diagnose invasive infections caused by *Candida* species. These yeasts are commonly isolated from various body sites and secretions of subjects who have no evidence of invasive disease (*2*). Yet many patients with invasive candidiasis do not have positive blood cultures (*3*). Also, recovery of the yeast from positive blood cultures is often delayed (*4*). Because a positive blood culture may not reflect invasive disease, patients may recover from transient fungemia without specific antifungal therapy. Similarly, detection of agglutinating and precipitating antibodies does not correlate consistently with invasive disease, especially in heavily immunosuppressed patients whose immunoglobulin function or levels may be depressed (*5*).

An alternative to standard culture and serologic diagnosis of invasive fungal infections involves chemical or immunologic detection of fungal cell metabolites. For example, the latex agglutination test for cryptococcal capsular polysaccharide is used widely to diagnose cryptococcal infection by the detection of antigen in blood and cerebrospinal fluid (*6*). Crossed electrophoresis (*7*), enzyme-linked immunosorbent assay (*8*), and passive hemagglutination inhibition (*9*) have been used to diagnose invasive candidiasis by detecting the circulating antigen. These reports suggest that antigen can be detected in the serum of some pa-

0036-8075/79/1102-0577$00.75/0