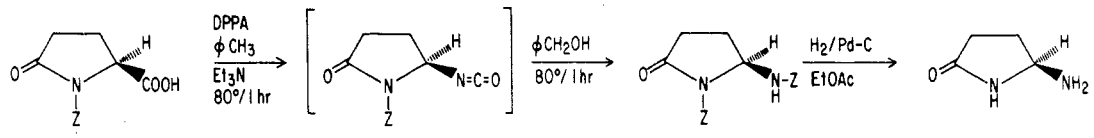
Scheme II



14

analogues was published recently.[39]

Concluding Remarks

Although the utilization of the topological approach to the synthesis of analogues of biologically active peptides is not new, the application of the conceptual approach to linear peptides has not been presented previously in a systematic manner. We have tried to place the terminology on a sound footing. We believe that the new term, "end group modified retro-inverso-peptide", defines precisely the types of transformations to which a linear peptide is subjected. We clearly stress that the modified peptides retain a full topological relationship to the parent peptide. Our studies on partially modified retro-inverso-peptides is a novel approach to the preparation of modified peptides through which we hope to assess the relative importance of side-chain vs. backbone structure for biological activity. Enhanced resistance to biodegradation processes might be another benefit of this type of structural modification.

Since this field is in its early stages, more examples of suitable analogues of biologically active peptides are needed to have a better understanding of the scope and limitations of the approach. Spectroscopic and theoretical conformational analyses will provide insights into the effects of such modifications on conformational preferences. With this information we should be better able to choose appropriate analogues of biologically active peptides to be synthesized and studied.

This study has been supported by the National Institutes of Health through Grants AM 15410-08 and FD 00590-04. We are also grateful to Mr. Wayne Becktel for his most helpful discussions and insight.

(39) M. Chorev, C. G. Willson, and M. Goodman, *J. Am. Chem. Soc.*, **99**, 8075 (1977).

Interactions in Aqueous Solution

HAROLD A. SCHERAGA

Department of Chemistry, Cornell University, Ithaca, New York 14853

Received May 22, 1978

The behavior of aqueous solutions of small and large molecules is influenced considerably by the nature of the solvent itself.[1-6] Thus, water plays a very important role in determining the properties of colloidal and macromolecular systems and, in particular, the manner in which proteins acquire their native structure and then interact with other small and large molecules.[7] In the absence of water, the interactions between the various functional groups of a polypeptide chain can be described in terms of empirical potential energy functions that have been parameterized with crystal-structure and gas-phase data on small molecules.[8] However, since protein folding occurs in water, the final conformation is influenced strongly by the solvent. Therefore, it is necessary to understand the nature of the interactions between water and the functional groups of proteins.

Several different types of approaches have been taken to investigate such interactions in aqueous solutions.[9] Initially, these involved the formulation of suitable models and their treatment by statistical mechanical methods. The cluster model of Nemethy and Scheraga,[10-13] with its improvements by Hagler et al.,[14] Lentz

Harold A. Scheraga was born in Brooklyn, New York. He attended the City College of New York, where he received his B.S. degree, and went on to graduate work at Duke University, receiving the Ph.D. in 1946, and, in 1961, a Sc.D. (Hon). Following postdoctoral work at Harvard Medical School, he joined the faculty at Cornell University, where he is Todd Professor of Chemistry. His research interests are in the physical chemistry of proteins and other macromolecules, chemistry of blood clotting, and structure of water and dilute aqueous solutions. This Account is based on Professor Scheraga's Award address for the 1978 ACS Award in Colloid or Surface Chemistry sponsored by Kendall Co.

(1) W. Kauzmann, *Adv. Protein Chem.*, **14**, 1 (1959).
(2) D. Eisenberg and W. Kauzmann, "The Structure and Properties of Water", Oxford University Press, Oxford, 1969.
(3) R. A. Horne, "Water and Aqueous Solutions", Wiley-Interscience, New York, N.Y., 1972.
(4) F. Franks, "Water, A Comprehensive Treatise", Plenum Press, New York, N.Y.: Vol. 1, 1972; Vol. 2, 1973; Vol. 3, 1973; Vol. 4, 1975; Vol. 5, 1975.
(5) W. A. P. Luck, "Structure of Water and Aqueous Solutions", Verlag Chemie, Weinheim, Germany, 1974.
(6) A. Ben-Naim, "Water and Aqueous Solutions", Plenum Press, New York, N.Y., 1974.
(7) G. Nemethy and H. A. Scheraga, *Q. Rev. Biophys.*, **10**, 239 (1977).
(8) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).
(9) H. A. Scheraga, *Ann. N.Y. Acad. Sci.*, **303**, 2 (1977).
(10) G. Nemethy and H. A. Scheraga, *J. Chem. Phys.*, **36**, 3382 (1962).
(11) G. Nemethy and H. A. Scheraga, *J. Chem. Phys.*, **36**, 3401 (1962).
(12) G. Nemethy and H. A. Scheraga, *J. Phys. Chem.*, **66**, 1773 (1962).
(13) G. Nemethy and H. A. Scheraga, *J. Chem. Phys.*, **41**, 680 (1964).
(14) A. T. Hagler, H. A. Scheraga, and G. Nemethy, *J. Phys. Chem.*, **76**, 3229 (1972).

0001-4842/79/0112-0007$01.00/0

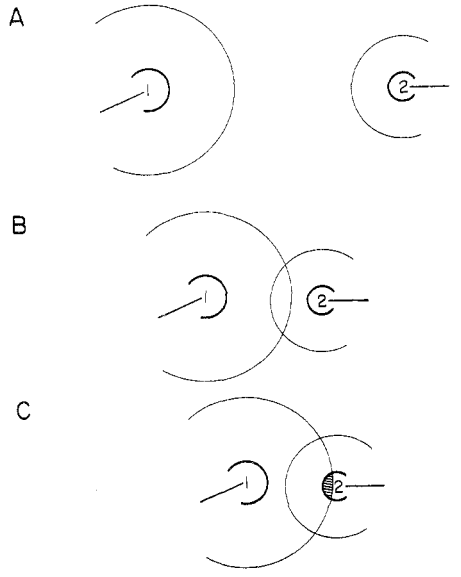


Figure 1. Schematic diagram illustrating three stages of the approach of two groups, 1 and 2. (A) No overlaps. (B) Overlap of the hydration spheres of groups 1 and 2; the free energy of hydration of both groups is unaltered. (C) Overlap of the van der Waals sphere of group 2 with the hydration sphere of group 1; the free energy of hydration of group 2 is unaltered, but that of group 1 is changed because of removal of water in the shaded region of the hydration shell of group 1. When groups 1 and 2 approach even more closely, the free energy of hydration of group 2 would change as well.[35]

et al.,[15] and Owicki et al.,[16] is an example of this approach. More recently, the properties of aqueous solutions have been derived from empirical potential-energy functions, solving integral equations for observable quantities,[17-19] as in the general theory of liquids, or by carrying out molecular dynamics[20-24] or Monte Carlo calculations.[25-34]

Potentially, each of these approaches can account for the behavior or reactivity of both nonpolar and polar functional groups in water. The role of hydration can be either nonspecific (expressed in terms of the presence of a solvent shell around the solute), as in the case of hydrophobic bonding between nonpolar groups, or specific, as in the case of polar groups where, in addition

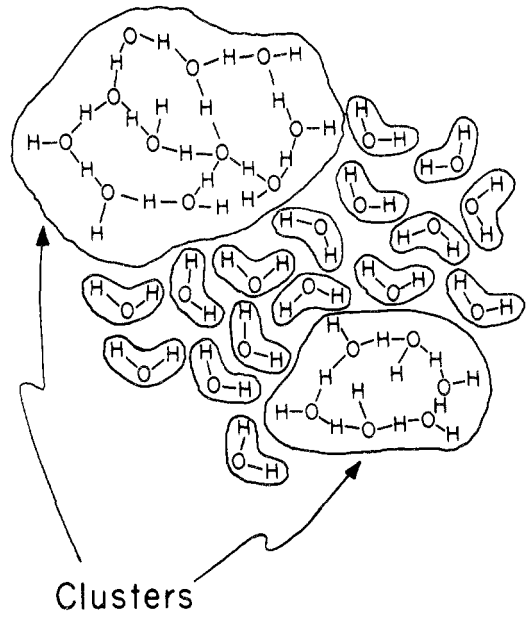


Figure 2. Schematic representation of liquid water in the Nemethy–Scheraga model, showing hydrogen-bonded clusters and unbonded molecules. The molecules in the interior of the clusters are tetracoordinated, but not drawn as such in this two-dimensional diagram.[10]

to nonspecific hydration, there is a specific interaction in which a water molecule competes for a hydrogen-bonding site on the solute molecule.[35]

In conformational energy calculations on proteins, a shell model is used to treat hydration.[35-37] In this model (Figure 1), a layer of water, having the thickness of one water molecule, is assumed to exist around each functional group. As the conformation of the protein is altered (to minimize its conformational energy), water must be eliminated from the hydration layers whenever they overlap in the manner indicated in Figure 1. The free-energy change accompanying this "dehydration" depends on the amount of water eliminated from the first layers (which, in the computations, depends on the degree of overlap) and on the free energy of hydration of the various groups. Such a model leads to a tendency for polar groups to lie on the surface of proteins, in contact with water, and for nonpolar groups to lie in the nonpolar interior, out of contact with water. The free energies of hydration of the various functional groups were estimated[36,37] from a variety of physical chemical data, but experience with these computations has demonstrated a need to obtain more accurate values of the free energies of hydration.

For this purpose, considerable attention has been focused on the structure and thermodynamic properties of water and of aqueous solutions of nonpolar and polar solutes, which can serve as models for understanding the hydration of the functional groups of proteins. We therefore shall consider some of the early and more recent developments in the theory of water and aqueous solutions. It is hoped that such studies will provide a more accurate description of hydration that can then be used to improve the shell model, whose simplicity

(15) B. R. Lentz, A. T. Hagler, and H. A. Scheraga, *J. Phys. Chem.*, **78**, 1531 (1974).
(16) J. C. Owicki, B. R. Lentz, A. T. Hagler, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2352 (1975).
(17) H. C. Andersen, *J. Chem. Phys.*, **61**, 4985 (1974).
(18) L. J. Lowden and D. Chandler, *J. Chem. Phys.*, **61**, 5228 (1974).
(19) L. R. Pratt and D. Chandler, *J. Chem. Phys.*, **67**, 3683 (1977).
(20) A. Rahman and F. H. Stillinger, *J. Chem. Phys.*, **55**, 3336 (1971).
(21) F. H. Stillinger and A. Rahman, *J. Chem. Phys.*, **57**, 1281 (1972); **60**, 1545 (1974); **61**, 4973 (1974).
(22) A. Rahman and F. H. Stillinger, *J. Am. Chem. Soc.*, **95**, 7943 (1973).
(23) A. Rahman, F. H. Stillinger, and H. L. Lemberg, *J. Chem. Phys.*, **63**, 5223 (1975).
(24) A. Geiger, A. Rahman, and F. H. Stillinger, *J. Chem. Phys.*, in press.
(25) J. A. Barker and R. O. Watts, *Chem. Phys. Lett.*, **3**, 144 (1969).
(26) R. O. Watts, *Mol. Phys.*, **28**, 1069 (1974).
(27) G. N. Sarkisov, V. G. Dashevsky, and G. G. Malenkov, *Mol. Phys.*, **27**, 1249 (1974).
(28) G. C. Lie and E. Clementi, *J. Chem. Phys.*, **62**, 2195 (1975).
(29) G. C. Lie, E. Clementi, and M. Yoshimine, *J. Chem. Phys.*, **64**, 2314 (1976).
(30) J. C. Owicki and H. A. Scheraga, *J. Am. Chem. Soc.*, **99**, 7403 (1977).
(31) J. C. Owicki and H. A. Scheraga, *J. Am. Chem. Soc.*, **99**, 7413 (1977).
(32) J. C. Owicki and H. A. Scheraga, *Chem. Phys. Lett.*, **47**, 600 (1977).
(33) J. C. Owicki and H. A. Scheraga, *J. Phys. Chem.*, **82**, 1257 (1978).
(34) S. Swaminathan and D. L. Beveridge, *J. Am. Chem. Soc.*, **99**, 8392 (1977).
(35) Z. I. Hodes, G. Nemethy, and H. A. Scheraga, *Biopolymers*, submitted for publication.
(36) K. D. Gibson and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **58**, 420 (1967).
(37) A. J. Hopfinger, *Macromolecules*, **4**, 731 (1971).

is ideal for conformational energy calculations on proteins.

Example of an Approach Based on a Model, and Its Implications

Nemethy and Scheraga[10,11] treated a model of water and aqueous solutions of hydrocarbons that was based on ideas of Frank and Wen.[38,39] The properties of the hydrophobic bond were derived directly from the properties of hydrocarbon solutions.[12]

The Frank–Wen model of water was based on the concept that the formation of hydrogen bonds in the liquid is a cooperative phenomenon, i.e., the bonds are not made and broken singly but several at a time, thus producing short-lived, compact, nearly spherical "clusters" of highly hydrogen-bonded regions surrounded by nonhydrogen-bonded molecules (Figure 2). A statistical mechanical treatment of this model gave the mean cluster size and the thermodynamic parameters of H_2O[10] and D_2O[13] at various temperatures.

The properties of hydrocarbon solutions were treated by focusing attention on the thermodynamic parameters for the process in which 1 mol of hydrocarbon is transferred from a nonpolar solvent into a dilute aqueous solution.[1] These solutions are nonideal in that the changes in volume and enthalpy are negative and there is a very large negative excess entropy change over the entropy of ideal mixing; this leads to a large positive change in free energy and hence to a low solubility because of the dominance of the entropy term over the enthalpy one.[1,40] Following Frank and Evans,[40] these changes were considered to arise from an ordering of water due to an increase in the degree of hydrogen bonding;[11] this behavior was attributed[11] to partial cage (or clathrate) formation, due to the interaction of the hydrocarbon with the water in the first solvation shell around the solute molecule which increased the fraction of tetracoordinated or "ice-like" molecules. The decrease in volume was attributed to the efficient filling of space in the partial cages that form near the solute. Since the number of water molecules in the first solvation shell increases with hydrocarbon size, there is a monotonic variation of the thermodynamic parameters for mixing with increase in molecular size of the hydrocarbons.

The theory developed to treat aqueous hydrocarbon solutions[11] made use of adjustable parameters to match the experimental thermodynamic properties. Thus, in essence, the theory for such solutions is independent of the details of the theory for pure liquid water. Hence, modifications of the theory for liquid water[14–16] would not be expected to modify the thermodynamic parameters for aqueous hydrocarbon solutions or for the formation of hydrophobic bonds derived therefrom.

Since the association of hydrocarbons in water annihilates some hydrocarbon–water contacts, the association process is analogous to that in which hydrocarbon is removed from solution. Thus, the thermodynamic parameters for formation of hydrophobic bonds have opposite signs from those for mixing hydrocarbon and water, viz., $\Delta H° > 0$, $\Delta V° > 0$, $\Delta S° > 0$, and $\Delta G° < 0$. Also, these parameters have similar

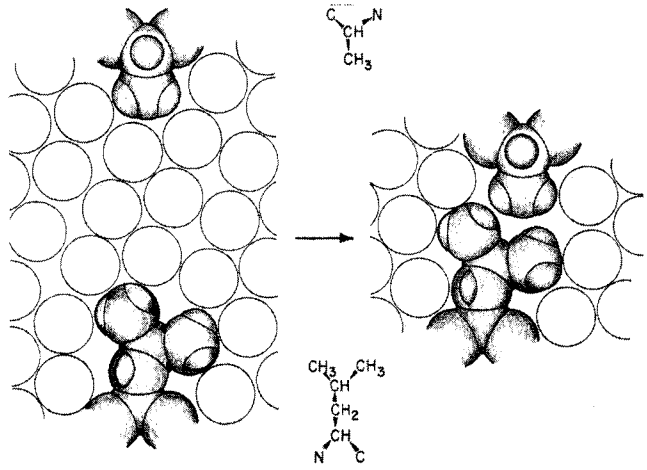


Figure 3. Schematic representation of the formation of a hydrophobic bond between two isolated side chains (alanine and leucine). The "bond" is formed through an approach of the two side chains until they touch, with a reduction of the number of nearest water neighbors. Water molecules are shown only schematically, without indicating particular orientations or hydrogen-bonded networks.[12]

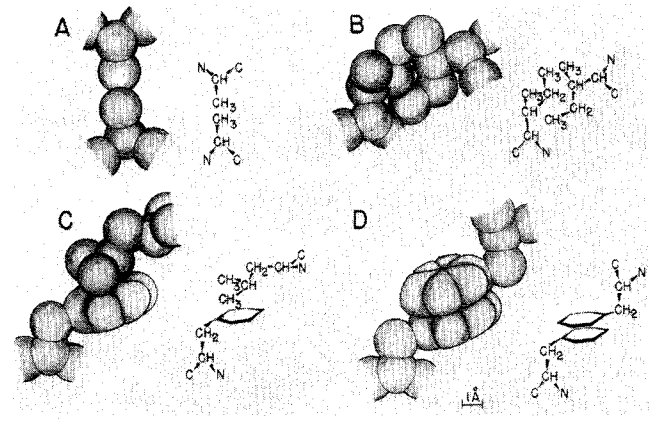


Figure 4. Illustrative examples of hydrophobic bonds between pairs of isolated side chains. The hydrogens are not indicated individually. Drawn to scale, but with the van der Waals radii reduced by 20% for the sake of clarity. The structural formulas to the right of each space-filling drawing indicate the arrangements of the atoms. (A) Alanine···alanine. (B) Isoleucine···isoleucine. (C) Phenylalanine···leucine. (D) Phenylalanine···phenylalanine.[12]

Table I
Theoretical Thermodynamic Parameters for Formation of the Hydrophobic Bonds of Figure 4 at 25 °C[11]

side chains	$\Delta G°$, kcal/mol	$\Delta H°$, kcal/mol	$\Delta S°$, eu
alanine· · ·alanine	−0.3	0.4	2.1
isoleucine· · ·isoleucine	−1.5	1.8	11.1
phenylalanine· · ·leucine	−0.4	0.9	4.7
phenylalanine· · ·phenylalanine	−1.4	0.8	7.5

temperature dependences as those for the mixing process. The formation of hydrophobic bonds can be considered to be accompanied by a *disordering* of water (i.e., partial melting of the "ice-like" regions that were produced when the separate hydrocarbons were first introduced into water). The formation of a hydrophobic bond is represented schematically in Figure 3, and some specific hydrophobic bonds are illustrated in Figure 4. The thermodynamic parameters for formation of the hydrophobic bonds of Figure 4 are given in Table I.[11] Experimental verification of these parameters has been

(38) H. S. Frank and W. Y. Wen, *Discuss. Faraday Soc.*, **24**, 133 (1957).
(39) H. S. Frank, *Proc. R. Soc. London, Ser. A*, **247**, 481 (1958).
(40) H. S. Frank and M. W. Evans, *J. Chem. Phys.*, **13**, 507 (1945).

discussed elsewhere.[41] Many other laboratories have contributed to our understanding of hydrophobic bonding, and their work is cited in ref 9. All of these approaches to hydrophobic bonding have similar objectives, are based on similar kinds of experimental data, apply similar models to account for the observations, and are all in agreement on the essential features of the phenomena involved. More recent studies indicate that the hydrophobic interaction includes not only the configurations indicated in Figures 3 and 4, but also one in which a layer of water exists between the nonpolar partners;[19,24] such a configuration had previously been considered to be less stable than the ones in Figures 3 and 4.[12]

Parameters of the magnitude of those of Table I account for the properties of many aqueous systems,[41] e.g., dimerization of aliphatic carboxylic acids in water, hydrophobic chromatography, stabilization of the helical conformation of polyamino acids containing nonpolar side chains, protein–protein association, etc. These parameters also account for the properties of detergent micelles, viz., the effect of solvent structure on the critical micelle concentration and the variation of the most probable micelle size with temperature and concentration in nonionic[42] and ionic[43] systems.

These parameters have also been used in an Ising model treatment of pockets in random-coil conformations of nonpolar polyamino acids in water,[44] and this concept has been extended recently[45] to develop a method for predicting nucleation sites for protein folding. The amino acid sequence of the protein is searched for pockets of nonpolar residues (see Figure 5) whose (negative) free energy of interaction compensates for the increase in free energy that is required to bring them into contact to form hydrophobic pockets; the pocket of lowest free energy is the predicted initial nucleation site. The predicted nucleation sites (and their associated electrostatic properties) have been used to rationalize some equilibrium and kinetic results on protein folding, including the relative amplitudes of absorption by transient species observed in kinetic studies. For example, the predicted nucleation site at residues 106–118 in bovine pancreatic ribonuclease is in agreement with an observed bend at residues 113–114[46] and with immunological experiments on the folding of this protein.[47] It also accounts for the biphasic kinetics observed in the refolding of the thermally unfolded protein.[48]

Using the aforementioned thermodynamic parameters (and related experimental data) for nonpolar groups in water and similar experimental data for polar groups in water, the shell model of Figure 1 has been parameterized.[35-37] By combining this shell model treatment of hydration with the interaction parameters obtained from crystal and gas-phase data,[8] the stable conformations of terminally blocked amino acids and

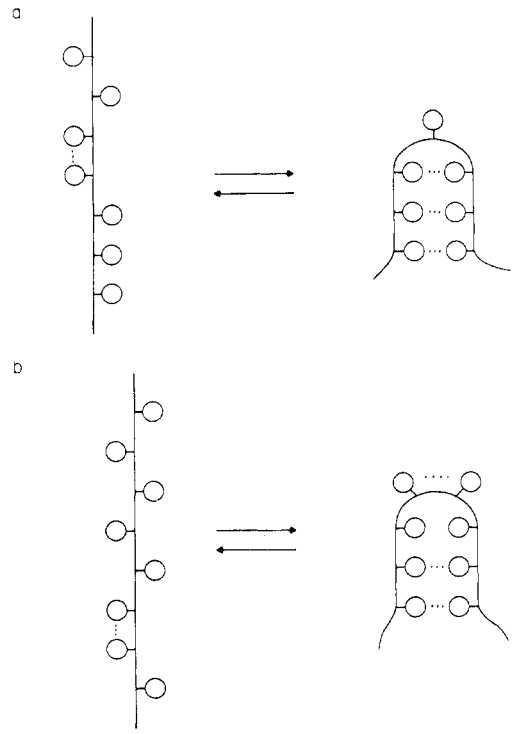


Figure 5. Schematic representation of nucleation step in which a hydrophobic pocket is formed from an ensemble of unfolded species [which *may* contain neighbor–neighbor hydrophobic bonds,[44] as indicated by the species on the left side of the equilibria in (a) and (b)]. In (a), only one side chain is involved in the "turn", and it does not participate in hydrophobic bonds with nearby residues. In (b), two side chains (which themselves can form a neighbor–neighbor hydrophobic bond) are involved in the "turn", and the pocket is shown as an imperfect one in that one pair of side chains is not sufficiently nonpolar to form a hydrophobic bond. Hydrophobic bonds are indicated by dotted lines. For pockets of types a and b, at least 5 and 4 residues, respectively, are assumed to be required, so that the pockets are large enough for the bends to be stereochemically feasible. Pockets of both types are included in the search algorithm to locate the nucleation site.[45]

dipeptides in water have been computed.[35] As a result, it has been possible to obtain conformational energy maps of unhydrated and hydrated *N*-acetyl-*N'*-methyl amino acid amides. Examination of these maps reveals a variety of effects due to hydration, e.g., a relative destabilization of the hydrogen-bonded C_7^{eq} ring conformation of *N*-acetyl-*N'*-methylalanine amide in the region $(\phi,\psi) = (-88°,79°)$, due to an opening of this ring and a lengthening of the N—H···O=C hydrogen bond. In addition, in *N*-acetyl-*N'*-methyl amides of dipeptides with polar side chains, the inclusion of hydration enhances the agreement between the probability of occurrence of β bends obtained from conformational energy calculations and that obtained from an examination of X-ray structures of proteins.[35]

The treatment of the Nemethy–Scheraga model of water has been improved,[14-16] with an attendant decrease in the mean size of clusters (which include hydrogen-bonded ring structures) (cf. Figures 2 and 6) and a reconciliation with continuum views of water structure. The distribution of cluster sizes shown in Figure 6 is similar to that obtained by Stillinger and Rahman[22] from a molecular dynamics treatment of liquid water. As indicated above, these modifications of the theory for liquid water would not be expected to modify the

(41) H. A. Scheraga, *Ann. N.Y. Acad. Sci.*, **125**, 253 (1965).
(42) D. C. Poland and H. A. Scheraga, *J. Phys. Chem.*, **69**, 2431 (1965).
(43) D. C. Poland and H. A. Scheraga, *J. Colloid Interface Sci.*, **21**, 273 (1966).
(44) D. C. Poland and H. A. Scheraga, *Biopolymers*, **3**, 315 (1965).
(45) R. R. Matheson, Jr., and H. A. Scheraga, *Macromolecules*, **11**, 819 (1978).
(46) H. W. Wyckoff, D. Tsernoglou, A. W. Hanson, J. R. Knox, B. Lee, and F. M. Richards, *J. Biol. Chem.*, **245**, 305 (1970).
(47) L. G. Chavez, Jr., and H. A. Scheraga, *Biochemistry*, **16**, 1849 (1977).
(48) P. J. Hagerman and R. L. Baldwin, *Biochemistry*, **15**, 1462 (1976).

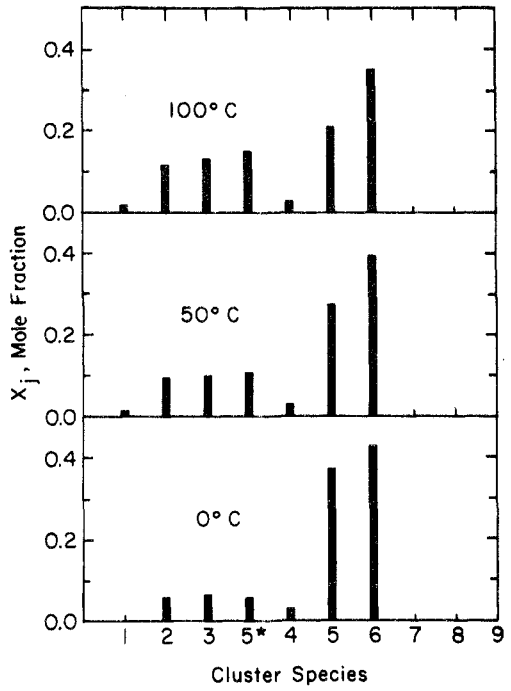


Figure 6. Calculated distributions of the mole fraction of water molecules in various cluster species at three temperatures.[15] See ref 15 for discussion of the species represented by 5*.

physical picture or thermodynamic parameters for aqueous hydrocarbon solutions and for the formation of hydrophobic bonds because they are based on parameterization with experimental data.

While some progress has been made with the model approach, and by application of a shell model of hydration based thereon to computations on peptides,[35] the need to obtain more accurate values of the free energies of hydration (in the shell model), especially for polar groups, has been recognized. Such thermodynamic quantities should be derivable directly from a potential-energy function properly parameterized on experimental data on aqueous solutions of small molecules. For example, it should be possible, with the methods discussed in the next section, to deduce the effect of hydration on the interaction of two molecules, A and B, in water, where both A and B can be either polar or nonpolar.

Recent Approaches

More recently, the theoretical studies of liquid water have abandoned the use of models and, instead, have used an empirical potential energy function together with statistical mechanical procedures to derive the properties of the system. The methods used were referred to in the introduction.[17-34]

Pair Potentials. In recent years, a number of empirical pair potentials have been used to treat water···water interactions. These include the Weissmann-Blum,[49] Ben Naim-Stillinger,[20,50] analytical fit to Hartree-Fock,[51,52] ST2,[22] Shipman-Scheraga,[53] and

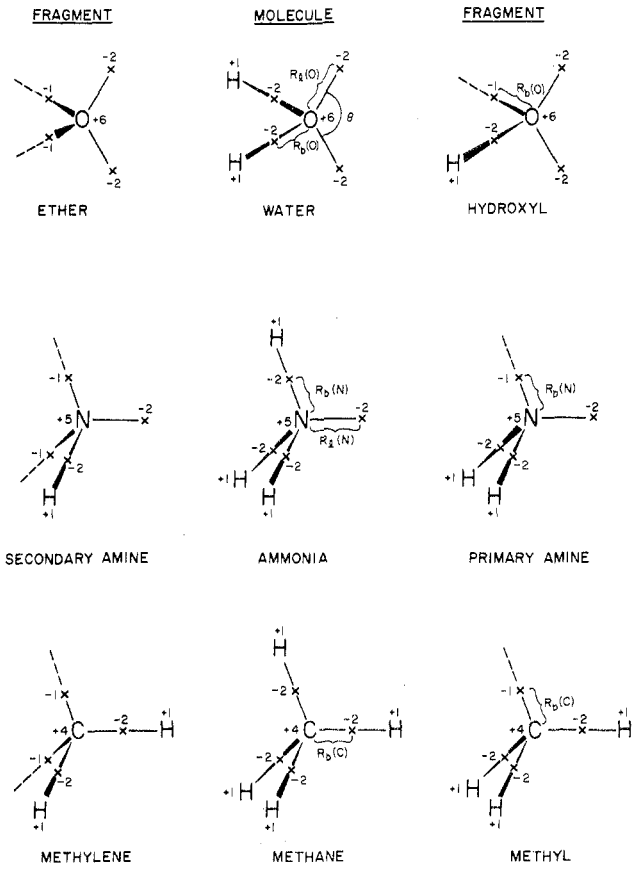


Figure 7. EPEN description of water, ammonia, and methane molecules together with two of their derivative fragments. The broken lines refer to the virtual bonds along which fragments are joined to assemble large molecules. The positions of the bonding and lone-pair electron point charges are indicated by ×'s. All charges are given in atomic units.[54]

EPEN (empirical potential using electrons and nuclei),[54-56] all of which were discussed by Owicki et al.,[57] and the configuration-interaction potential of Matsuoka et al.[58]

An advantage of the EPEN potential is that it is applicable not only to water but also to saturated and unsaturated organic molecules, which can be assembled from molecular fragments based on carbon, oxygen, and nitrogen atoms, respectively; the formation of molecules from molecular fragments (which are transferable, together with their parameters, from one molecule to another) is an essential feature of the EPEN approach. Figure 7 shows the EPEN formalism for the H_2O, NH_3, and CH_4 molecules, and the fragments derived therefrom. The EPEN potential has been applied to a variety of problems, e.g., to studies of the structure and energetics of crystals,[54,55] to calculations of dipole moments and barriers to internal rotation,[54-56] and to a study of the structure, energetics, and dynamics of small (hydrogen-bonded) water clusters.[57] In attempts

(49) M. Weissmann and L. Blum, *Trans. Faraday Soc.*, 64, 2605 (1968).
(50) A. Ben-Naim and F. H. Stillinger, Jr., in "Structure and Transport Processes in Water and Aqueous Solutions", R. A. Horne, Ed., Wiley, New York, N.Y., 1972, p 295.
(51) H. Popkie, H. Kistenmacher, and E. Clementi, *J. Chem. Phys.*, 59, 1325 (1973).
(52) H. Kistenmacher, G. C. Lie, H. Popkie, and E. Clementi, *J. Chem. Phys.*, 61, 546 (1974).
(53) L. L. Shipman and H. A. Scheraga, *J. Phys. Chem.*, 78, 909 (1974).
(54) L. L. Shipman, A. W. Burgess, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 543 (1975).
(55) A. W. Burgess, L. L. Shipman, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 854 (1975).
(56) A. W. Burgess, L. L. Shipman, R. A. Nemenoff, and H. A. Scheraga, *J. Am. Chem. Soc.*, 98, 23 (1976).
(57) J. C. Owicki, L. L. Shipman, and H. A. Scheraga, *J. Phys. Chem.*, 79, 1794 (1975).
(58) O. Matsuoka, E. Clementi, and M. Yoshimine, *J. Chem. Phys.*, 64, 1351 (1976).

to apply EPEN to Monte Carlo calculations on liquid water, it was found that nonhydrogen-bonded configurations of a water dimer were not sufficiently repulsive. The EPEN formalism was therefore modified[59,60] to rectify this deficiency and also to extend the original treatment of saturated molecules to unsaturated ones. Unlike potential functions that are parameterized to treat specific problems, EPEN has been parameterized on a large number of different types of physical properties.

As EPEN was used originally,[54-56] a molecule was constructed from molecular fragments consisting of a single heavy (nonhydrogen) atom nucleus, any hydrogen atom nuclei bonded to this nucleus, and point charges representing the bonding and lone-pair electrons associated with this heavy-atom fragment (see Figure 7). There were three types of interactions to be evaluated when using EPEN: (1) overlap repulsion between electrons, of the form $A \exp(-Br)$; (2) Coulombic interactions between all charge centers (electrons and nuclei), of the form $q_i q_j/r$, where q_i and q_j are the charges at two points separated by a distance r; and (3) dispersion and other attractive terms resulting from the interaction of fragments (this was taken to be effective between heavy atom nuclei only, and was of the form C/r^6). In the modifications that were introduced later,[59,60] the centers of the C/r^6-attractive terms were shifted from the heavy atom nuclei to the electrons, and a specific function for intramolecular rotations in unsaturated molecules where π electrons are involved was added.

An important aspect of EPEN is that it is parameterized on *experimental* data such as dipole moments, lattice constants, lattice energies, barriers to rotation, and so on. Such data provide information about a pair potential in the long-range attractive region and near the minimum. However, no experimental data are available to specify the shape of the potential in the repulsive region, and it is important to know the potential in the repulsive region, as well, for calculations on liquids. Therefore, resort has been had to ab initio calculations to determine the repulsive part of the EPEN potential. For water···water interactions, the potential of Matsuoka et al.[58] has been used.[59,60] For interactions of water with small molecules such as methane, methanol, and others, similar ab initio calculations are being carried out.[61] For the methane··· water interaction, Owicki and Scheraga[31] used an empirical fit to the ab initio results of Ungemach and Schaefer[62] to treat aqueous solutions of methane (see below). Preliminary results from the use of the modified EPEN potential in a Monte Carlo treatment of water[59,60] indicate that previous deficiencies in the repulsive region of the potential have been corrected; however, the modified EPEN potential does not reproduce the proper location of the second peak in the radial distribution curve of liquid water (even though it gives reasonable values of the thermodynamic parameters). An investigation is presently in progress to determine how the various features of a potential function influence the properties of liquids—in particular the radial distribution function. The Monte Carlo method provides a useful approach to explore this question. However, it should be pointed out that, in all Monte Carlo and molecular dynamics calculations carried out thus far, a pair potential (albeit an "effective" one) between water molecules is used, without taking three-body potentials into account.

Table II
Properties of Liquid Water Obtained by Monte Carlo Calculations[29,30]

	T,V,N	T,P,N	exptl
energy, kcal/mol	−6.5 (±0.3?)	−7.1 ± 0.3	−8.1
volume, cm^3/mol	(18.1)	23.8 ± 0.2	18.1
heat capacity, cal/mol deg	C_v = 13.5 (±2?)	C_p = 20.6 ± 7	18.0
compressibility, 10^{-6}/atm	53 (±?)	47 ± 16	41
coefficient of thermal expansion, 10^{-5}/deg	-	39 ± 13	27

Pratt and Chandler[19] treated aqueous solutions of hydrocarbons with a theory based on an integral equation for the pair correlation functions associated with spherical nonpolar species dissolved in water. They bypassed the problem involved in the theoretical computation of the properties of liquid water by using the experimentally determined oxygen···oxygen pair correlation function for the pure liquid. Thus, as in the Nemethy–Scheraga treatment of aqueous hydrocarbon solutions[11] and hydrophobic bonding,[12] the thermodynamic parameters for such solutions do not depend on a theoretical treatment of liquid water.

An empirical approach for treating long-range interactions (including solvent effects) in proteins has been developed by Tanaka and Scheraga.[63] It is based on the frequencies of contacts of pairs of amino acid residues in known crystal structures of a large number of proteins.

Monte Carlo Calculations on Liquid Water. Using molecular dynamics techniques, Rahman and Stillinger[20-24] have carried out a series of calculations on water and aqueous solutions, obtaining considerable insight into the properties of these systems. Similarly, interesting results have been obtained by Monte Carlo methods.[25-34]

Owicki and Scheraga[30] carried out a Monte Carlo simulation of liquid water in the isothermal–isobaric ensemble at 298 K and atmospheric pressure. Since the difficulties with the EPEN potential, referred to above, had not yet been rectified at the time that the Monte Carlo calculations were carried out,[30] the configuration-interaction potential of Matsuoka et al.[58] was used. Lie et al.[29] carried out similar computations in the isothermal–isovolumic ensemble at 298 K, with the volume fixed at its low-pressure value (for $P \sim 1$ atm). These results are compared in Table II; because of large errors in the computed values, the agreement is really not as satisfactory as it appears. The molar volume is overestimated in the T,P,N calculation; the value of the pressure in the T,V,N calculation was correspondingly in error. The preliminary calculations with the improved EPEN potential[59,60] gave better values of the

(59) J. Snir, R. A. Nemenoff, and H. A. Scheraga, *J. Phys. Chem.*, **82**, 2497 (1978).
(60) R. A. Nemenoff, J. Snir, and H. A. Scheraga, *J. Phys. Chem.*, **82**, 2504 (1978).
(61) D. H. Wertz, C. Tosi, E. Clementi, and H. A. Scheraga, *J. Chem. Phys.*, to be submitted.
(62) S. R. Ungemach and H. F. Schaefer III, *J. Am. Chem. Soc.*, **96**, 7898 (1974).
(63) S. Tanaka and H. A. Scheraga, *Macromolecules*, **9**, 945 (1976).

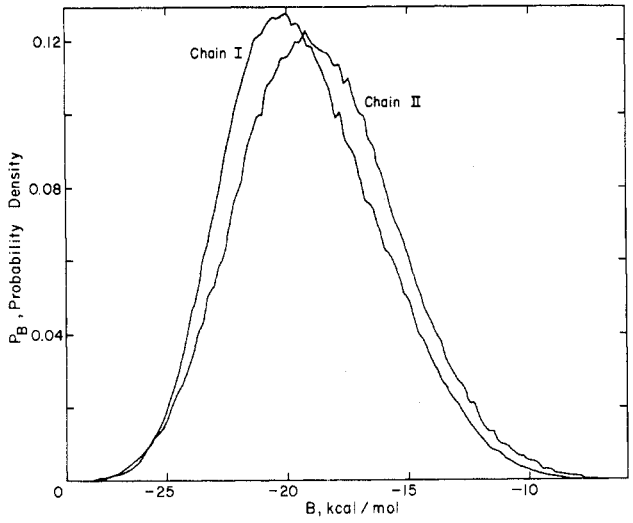


Figure 8. Probability distribution function, P_B, of binding energies B, where P_B describes the distribution of interaction energies of individual molecules with their surroundings. Chain I and chain II refer to calculations with 64 molecules (and simple cubic periodic boundary conditions) and with 100 molecules (and face-centered cubic periodic boundary conditions), respectively, in the Monte Carlo calculations. The mean binding energy is twice the mean potential energy per molecule.[30]

molar volume in the T,P,N ensemble but, as indicated above, the radial distribution curve was not reproduced well.

There are as yet serious practical difficulties in computing free energies by the Monte Carlo technique[30,31,33,64-66] which precluded the attempt to compute free energies in the work cited above. Whereas it is possible to compute the average potential energy, $\bar{U}$, efficiently in a Monte Carlo calculation, it is very difficult to obtain the average value of $\exp(U/kT)$, which is essentially the free energy. This difficulty arises because the probability distribution function for U has its maximum in the region where $\bar{U}$ appears, but it is close to zero in the region where large values of $\exp(U/kT)$ occur; thus, the Monte Carlo technique rarely samples the region of U where the value of $\exp(U/kT)$ is large. However, some attempts have been made[33] to circumvent this problem by using non-Boltzmann or umbrella sampling[64,65] in a computation of the free energy of a Lennard-Jones fluid containing a hard-sphere solute. Recently, Mezei et al.,[66] using a concept developed by Onsager[67] and Kirkwood,[68] carried out a series of Monte Carlo calculations of the free energy of liquid water by integrating with respect to a coupling parameter (see discussion in ref 33); making use of the potential function of Matsuoka et al.,[58] they obtained reasonable results for the free energy and entropy of liquid water.

Radial distribution functions and energy distribution functions were also computed by Owicki and Scheraga[30] and by Swaminathan and Beveridge,[34] an example of a distribution function being shown in Figure 8. It can be seen from Figure 8 that there is a broad, smooth distribution of binding energies in this model for liquid water.[30,34] Correspondingly, there is a smooth distribution of hydrogen-bond energy,[30] rather than relatively discrete sets of bonded and unbonded energies. Stillinger and Rahman[21] reached similar conclusions by a somewhat different route in their molecular dynamics study of a water model. Recently, using a modification of their earlier potential for water,[23,69] Stillinger and Rahman obtained an improved molecular dynamics representation of liquid water.[70]

Monte Carlo Treatment of Dilute Aqueous Methane. The Monte Carlo simulation described above was also carried out[31] for a dilute aqueous solution of methane in the isothermal–isobaric ensemble at 298 K and atmospheric pressure. Again, because the EPEN potential had not yet been reparameterized at the time that the Monte Carlo calculations were performed,[31] the potential used was that of Ungemach and Schaefer.[62]

Recognizing the large standard errors, the calculated partial molar energy and volume of the solute agree with the experimental values (a calculated partial molar energy of -11 ± 15 kcal/mol compared to an experimental value of -2.6 kcal/mol, and a calculated partial molar volume of 25 ± 34 cm^3/mol compared to an experimental value of 37 cm^3/mol). It was not possible to compute the partial molar free energy by the Monte Carlo technique because of the difficulty cited in the previous section.

Probably the most important single observation from these computations is that the first hydration shell is broad, extending from a C···O distance of ~3.1 to 6.0 Å, and contains ~23 water molecules. For comparison, in the computation on pure water, cited above, the first hydration shell extends from an O···O distance of ~2.4 to 3.6 Å and contains ~4 water molecules. Further, computed energy probability distribution functions indicate that the environments of water molecules in the first hydration shell around the methane molecule are characterized by lower, more sharply distributed energies than are the environments of water molecules far from the methane (i.e., in the bulk). These results support the concept[1,10-13,40] that nonpolar solutes increase the degree of hydrogen bonding or structure in their hydration shells. Such ordering is not strictly equivalent to the formation of a clathrate cage, but strong resemblances do exist—particularly between the calculated coordination number of methane (viz., ~23) and the sizes of the cages in the methane clathrate (viz., 20 and 24).[71] This ordering accounts for the thermodynamic properties of aqueous solutions of hydrocarbons,[1,11,40] and such systems had been used earlier as a model to compute the thermodynamic properties of the hydrophobic bond,[12,41] as described in the previous section.

This type of investigation, i.e., the use of pair interaction potential functions and Monte Carlo procedures to treat aqueous solutions, is being extended not only to improve the treatment of dilute aqueous solutions of methane but also to consider dilute solutions of polar molecules such as methanol,[61] methylamine, and so on, and to treat interactions between (polar and

(64) G. M. Torrie and J. P. Valleau, *Chem. Phys. Lett.*, **28**, 578 (1974).
(65) G. M. Torrie, Statistical Mechanics of Small Systems, Doctoral Thesis, University of Toronto, Toronto, Canada, 1975.
(66) M. Mezei, S. Swaminathan, and D. L. Beveridge, *J. Am. Chem. Soc.*, **100**, 3255 (1978).
(67) L. Onsager, *Chem. Rev.*, **13**, 73 (1933).
(68) J. G. Kirkwood, *J. Chem. Phys.*, **3**, 300 (1935).
(69) H. L. Lemberg and F. H. Stillinger, *J. Chem. Phys.*, **62**, 1677 (1975).
(70) F. H. Stillinger and A. Rahman, *J. Chem. Phys.*, **68**, 666 (1978).
(71) D. W. Davidson, ref 4, Vol. 2, Chapter 3.

nonpolar) solutes, in water, as indicated at the end of the previous section.

Concluding Remarks

The shell model,[36,37] with its recent improvements,[35] is a convenient one for including the effect of hydration in conformational energy calculations on proteins and in calculations of the interactions between macromolecules in aqueous solution. The parameters of this model can be improved by using the results of more accurate calculations on aqueous solutions of polar and nonpolar solutes. In the last analysis, the computations of the thermodynamic parameters for the interactions of such solutes in water must agree with experimental data. The theory can then be used to provide details of the interactions that are inaccessible to experiment; e.g., the free energy of hydration of a molecule can be partitioned into additive contributions from its component groups, for direct use in the shell model. The statistical mechanical methods described herein have the potential for yielding accurate theoretical data if a properly formulated potential function is used. The EPEN potential provides a framework, which, with improvements and modifications, can meet the requirements of a properly formulated potential function. Finally, the methods of umbrella sampling or integration with the use of a coupling parameter may help solve the problem of the computation of accurate free energies of hydration.

The work described here that was carried out in the author's laboratory was supported by research grants from the National Science Foundation (PCM75-08691) and the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (GM-14312).

Origin of the Pigments of Life: The Type-III Problem in Porphyrin Biosynthesis

ALAN R. BATTERSBY* and EDWARD MCDONALD

University Chemical Laboratory, Cambridge CB2 1EW, United Kingdom

Received February 2, 1978

Life on this planet depends ultimately on the photosynthetic activities of the chlorophylls,[1a] which are macrocyclic complexes of Mg^{2+}. In addition, electron transport in living systems and other vital functions such as reduction of oxygen and hydroxylation reactions are based on the cytochromes,[1b] which carry one or more units of a number of different heme systems (cf. **5**). When one adds that protoheme (**5**) is the oxygen-carrying unit of hemoglobin and myoglobin, it becomes evident that this whole family of macrocyclic pigments is of vital importance.

All the natural tetrapyrrolic macrocycles mentioned so far are derived in living things from uroporphyrinogen III (**1**), abbreviated throughout to uro'gen-III. In animals, plants, and most bacteria, the major pathway[2a] involves decarboxylation of **1** to yield copro'gen-III[2b] (**2**), which by oxidative decarboxylation produces proto'gen-IX (**3**). Aromatization to **4** followed by metal insertion yields protoheme (**5**), and, with esterification, the magnesium complex (**6**) for photosynthetic organisms. The intermediates **5** and **6** are then used to build one set of life's pigments, as indicated in Scheme I.

A second recently discovered pathway operates in certain bacteria, yeasts, and spinach to convert uro'gen-III (**1**) into siroheme (**8**), the prosthetic group of several sulfite reductases and nitrite reductases. Presumably siroheme (**8**) is derived from sirohydrochlorin[3] (**7**), and this substance (or possibly a dihydro form of it) has been proved[3b] to be a precursor of cobyrinic acid (**9**) and so of coenzyme B_{12}. Scheme II shows the relationships of this second set of pigments.

The central role of uro'gen-III (**1**) as the parent of all these vitally important metalloporphyrins, -chlorins, -isobacteriochlorins, and -corrins highlights the problem of its biosynthesis, a problem which is fascinating in its own right.

Enzymes and Building Blocks for Uro'gen-III. The pioneering studies of Shemin, Granick, Bogorad, Neuberger, and Rimington[2] established that two enzymes are required to catalyze the conversion of 4 mol of the monopyrrole, porphobilinogen, PBG (**10**), into uro'gen-III (**1**) and ammonia (Scheme III). The names of these enzymes, PBG-deaminase and uro'gen-III

A. R. Battersby graduated at the University of Manchester and received the Ph.D. and D.Sc. degrees from the Universities of St. Andrews and Bristol. He spent a period in the United States at the Rockefeller Institute for Medical Research and at the University of Illinois. In 1962, he became Professor of Organic Chemistry at the University of Liverpool where he remained until 1969 when he moved to Cambridge. He is a Fellow of the Royal Society.

Edward McDonald was born in Ormskirk, Lancashire, in 1943, and is University Lecturer at the University of Cambridge. During 1967–1969, he worked as a postdoctoral fellow with E. J. Corey at Harvard, after graduating from Imperial College and obtaining his Ph.D. from the University of Liverpool. Besides his research interest in the chemistry of porphyrins and corrins, he works on phenol oxidation and the synthesis of natural products.

(1) Reviewed in (a) "The Chlorophylls", L. P. Vernon and G. R. Seely, Eds., Academic Press, New York, 1966; (b) R. Lemberg and J. Barrett, "Cytochromes", Academic Press, London, 1973.

(2) (a) Reviewed by A. R. Battersby and E. McDonald, "Porphyrins and Metalloporphyrins", K. M. Smith, Ed., Elsevier, Amsterdam, 1975, p 61. (b) Abbreviations used: PBG, porphobilinogen; ALA, δ-aminolevulinic acid; uro'gen-III, uroporphyrinogen III; copro'gen-III, coproporphyrinogen III; proto'gen IX, protoporphyrinogen IX.

(3) (a) Isolation: M. J. Murphy, L. M. Siegel, H. Kamin, and D. Rosenthal, *J. Biol. Chem.*, **248**, 2801 (1973); M. J. Murphy and L. M. Siegel, *ibid.*, **248**, 6911 (1973); and references therein; see also H. E. Jones and G. W. Skyring, *Aust. J. Biol. Sci.*, **27**, 7 (1974). (b) Structure: A. R. Battersby, K. Jones, E. McDonald, J. A. Robinson, and H. R. Morris, *Tetrahedron Lett.*, 2213 (1977); A. R. Battersby, E. McDonald, H. R. Morris, M. Thompson, D. C. Williams, V. Ya. Bykhovsky, N. I. Zaitseva, and V. N. Bukin, *ibid.*, 2217 (1977); A. R. Battersby, E. McDonald, M. Thompson, and V. Ya. Bykhovsky, *J. Chem. Soc., Chem. Commun.*, 150 (1978); R. Deeg, H.-P. Kriemler, H.-H. Bergmann, and G. Müller, *Z. Physiol. Chem.*, **358**, 339 (1977); A. I. Scott, A. J. Irwin, L. M. Siegel, and J. N. Shoolery, *J. Am. Chem. Soc.*, **100**, 316 (1978); A. I. Scott, *Acc. Chem. Res.*, 11, 29 (1978).

0001-4842/79/0112-0014$01.00/0