Int. J. Biochem., 1971, **2**, 537–544. [Scientechnica (Publishers) Ltd.]

AN EMPIRICAL HYDROPHOBICITY SCALE FOR α-AMINO-ACIDS AND SOME OF ITS APPLICATIONS

AKINTOLA A. ABODERIN

Department of Biochemistry, College of Medicine,
University of Lagos, Lagos, Nigeria

(*Received* 17 *March*, 1971)

ABSTRACT

1. Hydrophobicity indices for α-amino-acids are derived from the relative mobilities of α-*N*-(4-nitrobenzofurazano)-amino-acids (NBD-amino-acids) on paper chromatograms with the solvent system ethyl acetate/pyridine/water, 8 : 2 : 1 (v/v).

2. The indices are algebraic sums of hydrophobicity contributions of functional groups, ΔH_f, present in the side chain.

3. Using the scale, amino-acid replacement matrices have been constructed and used to discuss aspects of the problem of structural homology in proteins.

A THOROUGH understanding of the mechanism of action of any protein requires an intimate knowledge of its three-dimensional structure, and an appreciation of the central role of the non-covalent interactions of amino-acid side chains which maintain that structure. To date, only X-ray crystallography, aided if need be by sequence analysis, can produce the necessary details of structure. However, even though most globular proteins are in principle crystallizable, not all proteins of interest have been or will be crystallized; these are, therefore, not susceptible to direct crystallographic analysis.

Current ideas on protein biosynthesis suggest the existence of indirect methods for arriving at the three-dimensional structures of polypeptides. Since linear polypeptides in non-denaturing media will spontaneously assume their native three-dimensional structure, a folding principle must *ipso facto* exist, by which it should be possible to derive the three-dimensional structure of a polypeptide given only its primary structure.

A folding principle must quantitatively reflect the contributions of two structural factors, size and polarity or hydrophobicity, to the interaction of amino-acid side chains. Of these two, steric factors are the easier to quantify. Molecular sizes or weights of amino-acid side chains can and have been used as steric indices (Epstein, 1967).

Attempts to arrive at numbers which reflect the relative polarity of the side chains of amino-acids have usually taken the form of determining the partition of a free amino-acid between aqueous and organic solvents. Two experimental approaches are solubility determinations (Tanford, 1962) and partition paper chromatography (Woese, Dugre, Saxinger, and Dugre, 1966). The use of free amino-acids in experiments designed to determine the polarity of the side chains is not ideal for two reasons:—

1. The effect of the polar α-amino and α-carboxyl groups cannot always be regarded as the same for all the amino-acids. Thus, for example, the pK_a values for the α-carboxyl groups of α-amino acids vary from 1·65 for cystine to 2·38 for tryptophan.

2. The α-amino and α-carboxyl groups do not exist as such in polypeptides except for the special cases of the *N*- and *C*-terminal amino-acids.

Another possible experimental approach which will take the above objections into consideration is to study the partition behaviour of suitable mono- or di-α derivatives of amino-acids. A suitable monosubstitution of α-amino-acids must be such as to lead to a

levelling of the differences in the relevant properties of the unsubstituted α-functional group.

The fluorescent amino-acid derivatives of 7 - chloro - 4 - nitrobenz - 2 - oxa - 1,3 - diazole (NBD-Cl) apparently satisfy this condition of suitability. α-*N*-NBD-Amino-acids have been observed to have identical values for the dissociation constant of the α-carboxyl group (Aboderin, 1971).

In this communication we propose an empirical hydrophobicity scale for the side chains of the α-amino-acids found in proteins, and discuss some of its applications. The scale has been constructed from the mobilities of α-*N*-NBD-amino-acids on 3 mm. Whatman chromatographic paper, using the monophasic apolar solvent ethyl acetate/pyridine/water, 8 : 2 : 1, as the mobile phase.

MATERIALS AND METHOD

Procedure for the synthesis and paper chromatographic separation of α-*N*-NBD-amino-acids has been described in an earlier communication (Aboderin and Kareem, 1971).

RESULTS AND DISCUSSION

1. The Hydrophobicity Scale

The R_F values of α-*N*-NBD-amino-acids on Whatman No. 3 paper, using the monophasic apolar solvent system ethyl acetate/pyridine/water (8 : 2 : 1), are presented in *Table I.* These are identical with earlier results except for the inclusion of the value for the histidine derivative and a revision of the arginine value.

It can be seen from these results that the common amino-acids found in proteins can be arranged in a unique sequence, which is a reflection of the polarities of the side chains of the amino-acids.* This observation has led to the construction of the proposed hydrophobicity scale which is shown in *Table II.* It has been obtained simply by setting the value for the most hydrophobic amino-acid—leucine—at 10 and computing the proportional values for the other amino-acids.

Table I.—R_F Values of α-*N*-NBD-Amino-acids*

Amino-acid	R_F
Asn	0·037
Asp	0·041
Lys	0·079
Gln	0·086
His	0·094
Glu	0·106
Arg	0·115
Ser	0·185
Thr	0·205
Gly	0·240
Pro	0·290
Ala	0·300
Tyr	0·470
Val	0·500
Met	0·510
Trp	0·540
Ile	0·550
Phe	0·562
Leu	0·590

* Values determined on Whatman No. 3 paper. Solvent: ethyl acetate/pyridine/water 8 : 2 : 1.

Characteristics of the Scale

a. Each amino-acid is characterized by a unique hydrophobicity index, H. This simple characteristic of the scale distinguishes it from existing scales (Tanford, 1962; Woese and others, 1966; Epstein, 1967). It affirms the analytic but important fact that no two amino-acids are the same, even though they may be very similar. The significance of this point in the evolution of the genetic code and of the protein biosynthetic machinery is obvious.

b. The difference in the indices of adjacent amino-acids $\Delta H \leqslant 1$. The exception to this rule is the large difference between Ala and Tyr for which $\Delta H = 2{\cdot}9$. This situation clearly permits one to distinguish between two groups of amino-acids:—

(i) Amino-acids with $H \leqslant 5{\cdot}0$.

(ii) Amino-acids with $H > 8{\cdot}0$.

* The remarkable paper chromatographic separations of sugars and their derivatives using this solvent system also seem to depend very strongly on the polarity of the compounds (Olaitan, 1961). Thus 6-deoxy-L-mannose (rhamnose) has a higher mobility than D-mannose and the bacterial didesoxyhexoses; e.g., paralose, abequose, and tyvelose move much faster than rhamnose. Similarly, deoxypentoses move much faster than the corresponding pentoses; and pentoses, being less polar, move faster than hexoses.

Those in the first group, except for those of neutral or intermediate polarity (Epstein, 1967), Ala, Pro, and Gly, are the so-called polar amino-acids, while those of the second group are the hydrophobic amino-acids.

Table II.—Empirical Hydrophobicity Indices for α-Amino-acids

Amino-acid	H
Asn	0·6
Asp	0·7
Lys	1·3
Gln	1·4
His	1·6
Glu	1·8
Arg	2·0
Ser	3·1
Thr	3·5
Gly	4·1
Pro	4·9
Ala	5·1
Tyr	8·0
Val	8·5
Met	8·7
Trp	9·2
Ile	9·3
Phe	9·6
Leu	10·0

c. The differences for homologous amino-acids which differ by a methylene group, ΔH_{CH_2}, are shown in *Table III*.

Table III.—Hydrophobicity Contribution of the Methylene Group

Amino-acids	ΔH_{CH_2}
Gln-Asn	0·8
Glu-Asp	1·1
Ala-Gly	1·0
Leu-Val	1·5
Average	1·1 ± 0·1

The relative constancy of these values suggests that the indices reflect the hydrophobicity of the side chains of amino-acids. This is analogous to the earlier results of Cohn and Edsall (1943) for the free energy of transfer of homologous amino-acids from water to non-polar solvents.

This result suggests that it should be possible to represent the hydrophobicity contribution of a given side chain (ΔH_s) as a sum of the hydrophobicity contributions of constituent chemical functional groups, ΔH_f, such as ΔH_{CH_2}. Since amino-acid side chains are made up of methylene and other functional groups, such as carboxamide, guanidine, etc., the hydrophobicity contribution of a functional group can be obtained as follows:—

We define a number ΔH_s for an amino-acid as

$$\Delta H_s = H_{\text{amino-acid}} - H_{\text{gly}}. \qquad (1)$$

It is clear that this number for any given amino-acid represents the hydrophobicity contribution of the side-chain characteristic of that amino-acid.

If, from the results shown in *Table III*, the hydrophobicity contribution of the methylene group (ΔH_{CH_2}) is taken as +1 (the polarity contribution will be −1), ΔH_f for a functional group, assuming additivity, can be obtained from the expression

$$\Delta H_f = \Delta H_s - n\Delta H_{CH_2} = \Delta H_s - n, \qquad (2)$$

where s represents the amino-acid side chain in which the functional group exists, and n is the number of methylene groups present.

ΔH_s and ΔH_{CH_2} thus defined are experimental numbers. To demonstrate the additive property of the indices it is necessary and sufficient to show that ΔH_f for any chemical functional group is independent of the compound from which it is calculated. There are two test cases involving the dicarboxylic amino-acids and their amides. Using equation (2) it can be shown that ΔH_{COOH} is −4·4 from Asp and −4·3 from Glu; and ΔH_{CONH_2} is −4·5 from Asn and −4·7 from Gln. The agreement within each set of numbers is satisfactory.

Values of ΔH_s and ΔH_f* for the various side chains and functional groups are shown in *Tables IV* and *V* respectively.

* Alternatively, polarity contributions of side chains and functional groups can be employed. If the polarity contribution of a side chain is defined as $\Delta H_s' = H_{gly} - H_{\text{amino-acid}}$ and equation (2) holds, the polarity contributions will be observed to be the negative of the hydrophobicity contributions.

The hydrophobicity contributions of the carboxylate and carboxamide groups (*Table V*), which reflect the relative hydrophobicities of the dicarboxylic amino-acids and their corresponding ω-amides, would seem to be at variance with commonly held views (*see* for example, Woese and others, 1966). However, measurements on the free energy change for the transfer of amino-acids from ethanol to water (Cohn and Edsall, 1943) demonstrate that the free energy change determined for the whole molecule and the side-chain contribution as calculated by Tanford (1962) are significantly more negative for the ω-carboxamides; i.e., the dicarboxylic amino-acids are less polar than their corresponding ω-carboxamides.

An analogous situation exists for the relative hydrophobicities of lysine and arginine.

This phenomenon can be most easily rationalized if it is assumed that the resonance-stabilized functional groups such as the carboxylate and the guanidinium groups are, as a result of the delocalization of charge, poorer hydrogen bond acceptors or donors than either a charge, such as the ammonium group, or strongly directed dipoles, such as the carbonyl group in carboxamides.

Table IV.—Hydrophobicity Contributions of Amino-acid Side Chains

Amino-acid	ΔH_s
Asn	−3·5
Asp	−3·4
Lys	−2·8
Gln	−2·7
His	−2·5
Glu	−2·3
Arg	−2·1
Ser	−1·0
Thr	−0·6
Gly	0·0
Pro	+0·8
Ala	+1·0
Tyr	+3·9
Val	+4·4
Met	+4·6
Trp	+5·1
Ile	+5·2
Phe	+5·5
Leu	+5·9

Table V.—Hydrophobicity Contributions of Functional Groups

Functional Group	H_r
Ammonium	−6·8
Guanidinium	−5·1
ω-Carboxamide	−4·6
ω-Carboxylate	−4·4
Imidazolium	−3·5
Hydroxyl: aliphatic; aromatic	−2·0; −1·6
Methylthioether	+0·6
Methylene, methyl	+1·0
Indole	+4·1
Phenyl	+4·5

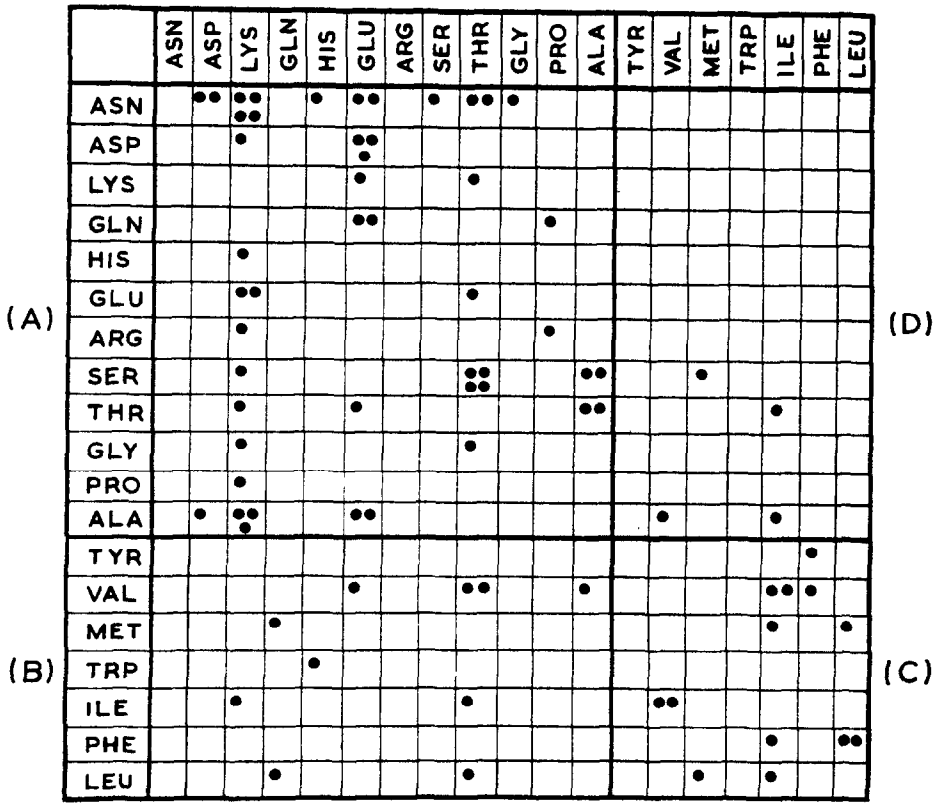
2. Amino-acid Replacement Matrix and its Uses

The availability of a unique ordering of amino-acids, based on the polarity of their side chains, permits the examination of questions relating to (1) ancestral homology in protein structure (Nolan and Margoliash, 1968), and (2) the role of side-chain polarity in the maintenance of tertiary structure, in a manner which seeks to make use of the informational content of both the identities and the differences in primary structural data of polypeptides.

Involved in this approach is the construction of what we term ' an amino-acid replacement matrix '. An example of such a matrix is given in *Fig.* 1, for which the extensive data on cytochrome *c* (Margoliash and Schejter, 1966) have been used. In the top row and left-hand column are the amino-acids arranged in order of increasing hydrophobicity. The top row is for the sequence of an arbitrarily chosen standard molecule, against which other homologous sequences are compared (left column). Thus, for example, if Lys in the standard sequence is replaced by Ser this will be scored as a point in the eighth row, third column. The replacement of an amino-acid at a given position in the primary structure by another amino-acid must be scored as 1 point only, regardless of the number of homologous molecules which are available for analysis. This must be so, since the mutation leading to the substitution of one amino-acid by another

could only have occurred once in evolutionary history.

The result is an array of points such as is shown in the figure. Utilizing the convenient division provided by the hydrophobicity scale, a replacement matrix can be divided into four unequal segments A, B, C, D, by segments B and D include replacements involving amino-acids of unlike polarities. These are respectively the conservative or permissible substitutions and the radical or non-permissible substitutions in the terminologies of Margoliash (1963) and Epstein (1964).



Standard sequence: Horse heart Cytochrome-C

FIG. 1.—Amino-acid replacement matrix for mammalian-type cytochrome *c*.

drawing dividing lines after Ala both vertically and horizontally. It should be borne in mind that, whilst the same distribution of points will exist in segments A and C when only two structures are compared, the results for segments B and D will be interchangeable, depending on which structure is chosen as standard.

The last qualification notwithstanding, the significance of such a partition of the replacement matrix is obvious. Segments A and C include those replacements involving amino-acids of similar polarities whereas those in

Amino-acid Replacement Matrix and Structural Homology

The assumptions which underlie any attempt to apply the replacement matrix approach to the question of homology in protein structure are (1) that the primary structure of a polypeptide determines the three-dimensional structure of that polypeptide, and (2) that non-covalent interactions between amino-acid side chains of different hydrophobicity are, more than any other, the most important factors in determining the tertiary structure of polypeptides. These two

assumptions are well founded experimentally (Goldberger, Epstein, and Anfinsen, 1963; Phillips, 1967). We must therefore expect some degree of regularity in replacement matrices, which should yield simple quantitative rules about homology in polypeptides, provided the hydrophobicity scale is realistic (*Table II*).

Data for the analysis of six different groups of homologous proteins: cytochrome *c* (Margoliash and Schejter, 1966), subtilisin (Smith, Delange, Evans, Landon, and Markland, 1968), bovine and rat pancreatic ribonuclease (Smyth, Stein, and Moore, 1963; Beintema and Gruber, 1967), bovine chymotrypsinogen A (Hartley, Brown, Kaufmann, and Smillie, 1965), bovine trypsinogen (Walsh and Neurath, 1964; Mikes, Holeysovsky, Tomasek, and Sorm, 1965), and chymotrypsinogen B (Smillie, Furka, Nagabhusan, Stevenson, and Parkes, 1968), as well as those for two groups of non-homologous proteins: staphylococcal nuclease (Taniuchi, Anfinsen, and Sodja, 1967) and sperm whale myoglobin (Edmundson, 1965); and ribonuclease T_1 (Takahashi, 1965) and human cytochrome *c* (Matsubara and Smith, 1963), are presented in *Table VI*.

These data show that the largest percentage of replacements is of the permissible polar type, i.e., those in segment A. This is, however, partly a reflection of the preponderance of polar amino-acids as defined above. Thus, even for the case of the two groups of structurally and functionally non-homologous proteins, 50 per cent of all replacements are found in this segment.

The identity of the results for segment A for the two groups of non-homologous proteins reveals that the replacement matrix approach offers a simple statistical test for structural homology. If two structurally dissimilar polypeptides of the same length are compared, a random distribution of permissible and non-permissible replacements must be expected; i.e., the ratio (A+C)/(B+D) must be unity, as the results for the non-homologous proteins show. Therefore, the departure from unity of this ratio is a measure of the non-randomness of amino-acid substitution. A more convenient number is the homology index which can be defined simply as (A+C)/(B+D) −1. This will then be zero for a complete lack of homology.

The homology indices for the proteins analysed in *Table VI* are summarized in

Table VI.—Amino-acid Replacement Matrix Analysis of Proteins

Proteins	Number of Molecules	Total No. of Substitutions	Percentage Total				(A+C)/(B+D)*
			A	B	C	D	
Mammalian-type cytochrome *c*	10	75	65	13	17	5	3·9
Pancreatic ribonuclease: bovine and rat	2	38	82	3	10	5	10·0
Subtilisin: BPN and Carlsberg	2	80	71	6	15	6	6·3
Bovine chymotrypsinogen A and B	2	50	68	14	12	6	3·5
Bovine chymotrypsinogen B and trypsinogen	2	129	58	10	17	15	2·6
Bovine chymotrypsinogen A and trypsinogen	2	114	57	13	16	14	2·4
Staphylococcal nuclease and sperm whale myoglobin	2	138	50	19	8	23	1·2
Ribonuclease T_1 and human cytochrome *c*	2	93	50	22	6	22	1·0

* Normalized by 1·15 which is the ratio $\frac{\text{No. of squares in A+C}}{\text{No. of squares in B+D}}$.

Table VII. That these results are in accord with those obtained using the method of Fitch (1966), which involves the calculation of minimal mutation distances between two structures, can be demonstrated with the chymotrypsinogen / trypsinogen system. Smillie and co-workers (1968) have concluded, on the basis of such calculations, that chymotrypsinogen A and B are more similar structurally to each other than is either protein to trypsinogen. The same conclusion is apparent from the present work.

Detection of Structural Homology

Implicit in these results is a simple, practical, and quantitative monitor for the detection of homology between two structures that are to be compared. A homology index is determined for each trial alinement of the two structures until a maximum value of the index is obtained; this must be the case if any homology exists, otherwise a random distribution of permissible and non-permissible replacements will always be obtained.

Cumulative Homology Indices

When more than two homologous molecules are compared, the resultant cumulative homology index is necessarily related to the extent to which the primordial gene has been conserved (in the case of divergent evolution) through its evolutionary history. We consequently have a measure for the relative degrees of conservation of different genes in phylogeny when the cumulative indices for different groups of homologous proteins are compared. Thus the gene coding for mammalian-type cytochrome *c*, with a cumulative index of 3 (derived from an analysis of 10 species), has been relatively better conserved than that coding for the α-chain of haemoglobin, with a cumulative index of 2 (derived from an analysis of 12 species). This conclusion is not at variance with what is to be expected from a consideration of the ecological pressures on the two groups of molecules throughout phylogeny.

CONCLUSION

The present approach to the question of structural homology has been based solely on the hydrophobicities of the side chains of the α-amino-acids. That it has yielded results, which are in accord with those obtained using other more involved approaches, serves to underline the realistic nature of the proposed hydrophobicity scale. More important, this approach helps to emphasize the point that mutations leading to viable homologous structures must be such as to lead to a conservation of the total hydrophobicity characteristic of the functional primordial molecule.

Table VII.—Homology Indices for Proteins

Protein	Homology Index
Cytochrome *c*	2·9
Pancreatic ribonuclease	9·0
Subtilisin	5·3
Chymotrypsinogen A and B	2·5
Chymotrypsinogen B and trypsinogen	1·6
Chymotrypsinogen A and trypsinogen	1·4
Staphylococcal nuclease and myoglobin	0·2
Ribonuclease T_1 and human cytochrome *c*	0·0

REFERENCES

Aboderin, A. (1971), unpublished results.

Aboderin, A., and Kareem, H. A. (1971), '7-Chloro-4-nitrobenz-2-oxa-1,3 diazole in protein chemistry. I. Preparation of L-amino acid derivatives', *J. West Afric. Sci. Ass.* (in the press).

Beintema, J. J., and Gruber, M. (1967), 'Amino acid sequence in rat pancreatic ribonuclease', *Biochem. biophys. Acta*, **147**, 612–614.

Cohn, E., and Edsall, J. T. (1943), *Proteins, Amino Acids and Peptides*, Chapter 9, pp. 196–216. New York: Reinhold.

EDMUNDSON, A. B. (1965), ' Amino acid sequence of sperm whale myoglobin ', *Nature, Lond.*, **205**, 883–887.

EPSTEIN, C. J. (1964), ' Relation of protein evolution to tertiary structure ', *Nature, Lond.*, **203**, 1350–1352.

EPSTEIN, C. J. (1967), ' Non-randomness of amino-acid changes in the evolution of homologous proteins ', *Nature, Lond.*, **215**, 355–359.

FITCH, W. M. (1966), ' An improved method of testing for evolutionary homology ', *J. mol. Biol.*, **16**, 9–16.

GOLDBERGER, R. F., EPSTEIN, C. J., and ANFINSEN, C. B. (1963), ' Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver ', *J. biol. Chem.*, **238**, 628.

HARTLEY, B. S., BROWN, J. R., KAUFMANN, D. L., and SMILLIE, L. B. (1965), ' Evolutionary similarities between pancreatic proteolytic enzymes ', *Nature, Lond.*, **207**, 1157–1159.

MARGOLIASH, E. (1963), ' Primary structure and evolution of cytochrome *c* ', *Proc. natn. Acad. Sci. U.S.A.*, **50**, 672–679.

MARGOLIASH, E., and SCHEJTER, A. (1966), ' Cytochrome *c* ', *Adv. Prot. Chem.*, **21**, 113–268.

MATSUBARA, H., and SMITH, E. L. (1963), ' Human heart cytochrome *c* ', *J. biol. Chem.*, **238**, 2732–2753.

MIKES, O., HOLEYSOVSKY, V., TOMASEK, V., and SORM, F. (1966), ' Covalent structure of bovine trypsinogen. The position of the remaining amides ', *Biochem. biophys. Res. Commun.*, **24**, 346–352.

NOLAN, C., and MARGOLIASH, E. (1968), ' Comparative aspects of primary structures of proteins ', *A. Rev. Biochem.*, **37**, 727–781.

OLAITAN, S. A. (1961), ' Polysaccharides of *Chlorella pyrenoidosa* ', Ph.D. Dissertation, Cambridge University, p. 144.

PHILLIPS, D. C. (1967), ' The egg white lysozyme molecule ', *Proc. natn. Acad. Sci. U.S.A.*, **57**, 484–495.

SMILLIE, L. B., FURKA, A., NAGABHUSAN, N., STEVENSON, K. J., and PARKES, C. O. (1968), ' Structure of chymotrypsinogen B compared with chymotrypsinogen B and trypsinogen ', *Nature, Lond.*, **218**, 343–346.

SMITH, E. L., DELANGE, R. J., EVANS, W. H., LANDON, M., and MARKLAND, F. S. (1968), ' Subtilisin Carlsberg. V. The complete sequence, comparison with subtilisin BPN; evolutionary relationships ', *J. biol. Chem.*, **243**, 2184–2191.

SMYTH, D. G., STEIN, W. H., and MOORE, S. (1963), ' The sequence of amino acid residues in bovine pancreatic ribonuclease. Revisions and confirmations ', *J. biol. Chem.*, **238**, 227–234.

TAKAHASHI, K. (1965), ' Amino acid sequence of ribonuclease T_1 ', *J. biol. Chem.*, **240**, 4117–4119.

TANFORD, C. (1962), ' Contribution of hydrophobic interactions to the stability of globular conformation of proteins ', *J. Am. chem. Soc.*, **84**, 4240–4247.

TANIUCHI, H., ANFINSEN, C. B., and SODJA, A. (1967), ' The amino acid sequence of an extracellular nuclease of *Staphylococcus aureus*. III. Complete amino acid sequence ', *J. biol. Chem.*, **242**, 4752–4758.

WALSH, K. A., and NEURATH, H. (1964), ' Trypsinogen and chymotrypsinogen as homologous proteins ', *Proc. Natn. Acad. Sci. U.S.A.*, **52**, 884–889.

WOESE, C. R., DUGRE, D. H., SAXINGER, W. C., and DUGRE, S. A. (1966), ' The molecular basis of the genetic code ', *Proc. natn. Acad. Sci. U.S.A.* **55**, 966–973.

Key Word Index: Amino-acids, side-chain hydrophobicity, structural homology, protein structure.