[CONTRIBUTION FROM THE DEPARTMENT OF BIOCHEMISTRY, UNIVERSITY OF WISCONSIN]

Water Solubility of Some Synthetic Polypeptides[1]

BY R. R. BECKER AND MARK A. STAHMANN

RECEIVED MAY 11, 1953

The effect of amino acid composition on the water solubility of several polypeptides was studied. It was found that when the N-carboxyamino acid anhydrides of glycine and L-leucine were polymerized independently only about 15% of the resulting products remained in solution. However, when these anhydrides were copolymerized, about 50% of the products were soluble in the aqueous reaction medium. The soluble products from the copolymerizations were of a higher molecular weight than those from independent polymerizations. Experiments in which γ-ethyl-L-glutamate anhydride was copolymerized with N-carboxyglycine anhydride gave similar results. The increased solubility of the copolymers as compared with polypeptides containing only one amino acid was interpreted as arising from a decrease in the amount of intermolecular hydrogen bonding. This decrease was attributed to steric hindrance produced by the side chains of the second amino acid. Infrared analysis supported this interpretation.

Ambrose and Elliott,[2-4] Bamford and associates,[5-7] and Pauling and Corey[8,9] have proposed structures for synthetic polypeptides that are similar in that they require complete hydrogen-bonding between all the carbonyl and imino groups of the peptide bonds, but differ in the manner in which the peptide chains are folded or coiled. Ambrose and Elliott have extended their studies to the transformation between the folded and extended forms, and have suggested that the relative proportions of the two forms may determine to a considerable extent the solubility properties of certain polypeptides and proteins. It was postulated that in the insoluble fibrous proteins the polypeptide chains were extended and intermolecularly hydrogen bonded to adjacent chains to a greater extent than in the soluble globular proteins.

Synthetic polypeptides are usually in the extended insoluble form with the peptide chains largely intermolecularly hydrogen bonded. It would be expected that the extent of this intermolecular hydrogen bonding in synthetic polypeptides would be reduced by steric interference if amino acids containing bulky side chains were introduced at random into the molecule. The aqueous polymerization of N-carboxyamino acid anhydrides[10] offered a convenient system to test this concept, since the polymers and copolymers could be readily prepared and the effects on solubility measured directly on the aqueous reaction mixtures.

In this study we prepared polypeptides containing only glycine, leucine or γ-ethyl-L-glutamic acid and compared their water solubility with polypeptides similarly prepared but containing two amino acids. In all cases the introduction of the second amino acid resulted in an increased water solubility of the resulting polypeptides, even when the second amino acid contained a large hydrophobic, aliphatic side chain.

(1) Published with the approval of the Director of the Wisconsin Agriculture Experiment Station. Supported in part by a research grant from the Herman Frasch Foundation.

(2) E. J. Ambrose and A. Elliott, *Proc. Roy. Soc.* (*London*), **A205**, 47 (1951).

(3) E. J. Ambrose and A. Elliott, *ibid.*, **A208**, 75 (1951).

(4) E. J. Ambrose and A. Elliott, *ibid.*, **A206**, 206 (1951).

(5) C. H. Bamford, W. E. Hanby and F. Happey, *ibid.*, **A205**, 30 1951).

(6) C. H. Bamford, *Proc. Roy. Soc. Med.*, **44**, 393 (1951).

(7) C. H. Bamford, W. E. Hanby and F. Happey, *Proc. Roy. Soc.* (*London*), **A206**, 407 (1951).

(8) L. Pauling, R. B. Corey and H. R. Branson, *Proc. Nat. Acad. Sci.*, **37**, 205 (1951).

(9) L. Pauling and R. B. Corey, *ibid.*, **37**, 235 (1951).

(10) R. R. Becker and M. A. Stahmann, *J. Biol. Chem.*, **204**, 737 (1953).

Experimental

The preparation of the N-carboxyamino acid anhydrides of glycine, L-leucine and γ-ethyl-L-glutamate, the polymerization techniques used, and the analyses for chain length were the same as already described.[10]

N-Carboxyamino acid anhydrides were polymerized both singly and together in phosphate buffer at *p*H 7.4. The insoluble polypeptides were separated by centrifugation, washed, lyophilized and weighed. This insoluble fraction and the soluble portion of the reaction mixtures were analyzed for total and α-amino nitrogen, and from these results the yield, molecular weight and chain length were calculated.

Results

Data from experiments in which 1% reaction mixtures of the N-carboxy anhydrides of glycine, L-leucine and γ-ethyl-L-glutamate were polymerized both independently and copolymerized are summarized in Table I. It should be noted that the reaction mixture which contained the anhydrides of glycine and γ-ethyl-L-glutamate in a molar ratio of 4 to 1 remained completely soluble, even though the average chain length was longer than that of any of the soluble products from independent polymerizations. Independent polymerization resulted in 74 and 49% insoluble polypeptide formation respectively from the glycine and γ-ethyl-L-glutamate anhydrides. The polypeptide containing all three amino acids showed about a 140% increase in solubility compared to polypeptides containing only glycine or leucine, and about 28% to those containing only γ-ethyl-L-glutamate

TABLE I

EFFECT OF COPOLYMERIZATION ON SOLUBILITY AND MOLECULAR WEIGHT

Reaction mixture contained 50 mg. of anhydride in 5 ml. of *M*/15 phosphate buffer, *p*H 7.4, 18 hours at 25°. The mole ratio of the copolymerizations is indicated by the figures preceding the anhydride

N-Carboxyamino acid anhydrides	% soluble	Chain length
Glycine	27	4.0
L-Leucine	26	2.3
γ-Et-L-glutamate	51	5.1
4,1-Co-glycine-γ-Et-L-glutamate	100	8.2
2,2,1-Co-glycine-L-leucine-γ-Et-L-glutamate	65	4.6

Table II summarizes data from experiments in which 2% reaction mixtures of the anhydrides of glycine and L-leucine were polymerized independently or copolymerized and both the soluble and in-

soluble fraction analyzed. When independently polymerized, only about 15% of the product remained in solution in each case. When 1-to-1 molar ratios of the anhydrides were copolymerized, 36% of the products were soluble. The soluble products from the copolymerization had almost twice the average molecular weight of those from the independent polymerizations.

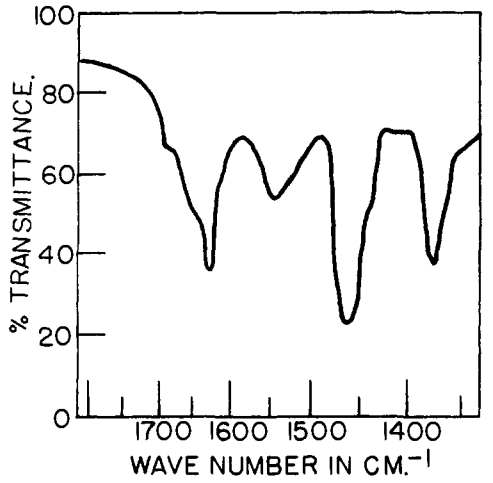

Fig. 1.—Infrared spectrum of polyglycine.

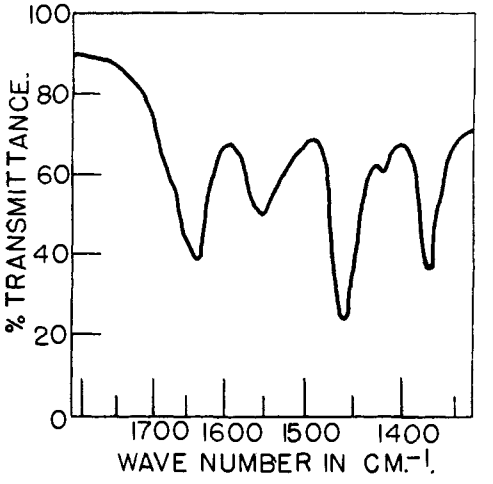

Fig. 2.—Infrared spectrum of poly-L-leucine.

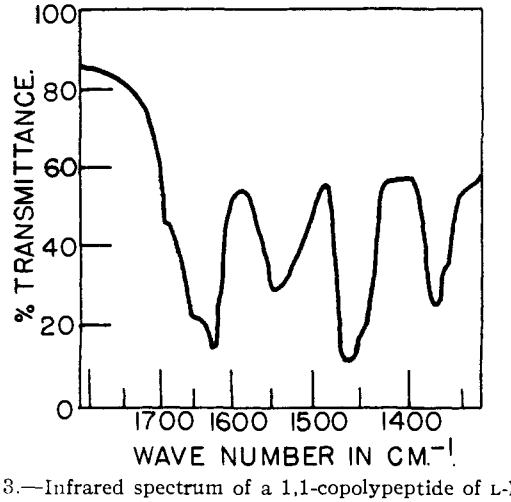

Fig. 3.—Infrared spectrum of a 1,1-copolypeptide of L-leucine and glycine.

TABLE II

EFFECT OF COPOLYMERIZATION ON SOLUBILITY AND MOLECULAR WEIGHT

Reaction mixture contained 500 mg. of anhydride in 25 ml. of $M/15$ phosphate buffer, pH 7.4, 48 hours at 25°

N-Carboxyamino acid anhydride	Insoluble Yield, %	Insoluble Mol. wt.	Insoluble Chain length	Soluble Yield, %	Soluble Mol. wt.	Soluble Chain length
Glycine	87	730	12.8	14	165	2.9
L-Leucine	90	950	6.8	16	180	1.6
1,1-Co-glycine-L-leucine	67	1260	14.9	36	350	4.1

To determine whether the increased solubility observed was in fact due to copolymerization, and not to such factors as solubilization of one peptide by another, experiments in which one amino acid was allowed to polymerize before the second was added were carried out. N-Carboxyglycine anhydride was polymerized for 7 hours, a time sufficient for complete polymerization of this anhydride,[10] and then either N-carboxy-L-leucine or N-carboxy-γ-ethyl-L-glutamate anhydride was added, and the reaction allowed to go to completion. This method should result in polypeptides containing a predominance of glycine on the carboxyl end of the polypeptide, and either glutamate or leucine on the amino end. Significant amounts of polypeptides containing only one amino acid may also be found. Of most importance, however, is the fact that a distribution of either leucine or γ-ethyl-glutamate residues along the polypeptide chain is excluded. The results in Table III show that no significant increase in solubility resulted from the delayed additions of anhydride as compared to the independent polymerizations. However, copolymerization resulted in an increase in the amount of soluble products from 13 to 49% for the co-glycine-L-leucine polypeptides and from 25 to 46% for the co-glycine-glutamate polypeptides.

TABLE III

EFFECT OF COPOLYMERIZATION *versus* INDEPENDENT POLYMERIZATION UPON SOLUBILITY AND MOLECULAR WEIGHT

Reaction mixtures contained 500 mg. of anhydride per 25 ml. of $M/15$ phosphate buffer, pH 7.4, 48 hours at 25°.

N-Carboxyamino acid anhydride	Insoluble Yield, %	Insoluble Mol. wt.	Insoluble Chain length	Soluble Yield, %	Soluble Mol. wt.	Soluble Chain length
Glycine	84	540	9.5	19	165	2.9
L-Leucine	94	965	8.6	11	170	1.5
γ-Et-L-glutamate	81	2580	17.0	21	470	3.0
1,1-Glycine-L-leucine[a]	88	1000	11.8	13	230	2.7
1,1-Co-glycine-L-leucine	57	1250	14.7	49	390	4.6
1,1-Glycine-γ-Et-L-glutamate[a]	74	1410	13.2	25	524	4.9
1,1-Co-glycine-γ-Et-L-glutamate	56	2270	21.2	46	556	5.2

[a] N-Carboxyglycine anhydride polymerized, followed in 7 hours by second anhydride.

The observation by Bamford and associates that the frequency of the CO stretching mode is often

slightly different in the folded (intramolecularly hydrogen bonded) than in the extended (intermolecularly hydrogen bonded) configurations of the polypeptide chain has been utilized to study the types of folding in polypeptide preparations.[5] Samples of polyglycine (average molecular weight 540), polyleucine (average molecular weight 960) and 1,1-co-glycine-L-leucine polypeptide (average molecular weight 1250) were dried at 55° and 0.1 mm., and their infrared absorption spectra determined in Nujol mull in the Baird Associates, Inc., I. R. Spectrophotometer, using a calcium fluoride prism.[11] Figures 1 to 3 give the spectra for these polypeptides. The carbonyl stretching mode for polyglycine occurs at about 1630 cm.$^{-1}$ (Fig. 1) in agreement with the published data of Ambrose and Elliott.[2] In poly-L-leucine, this peak occurs at about 1640 cm.$^{-1}$, with a shoulder at about 1660 cm.$^{-1}$ (Fig. 2). This shoulder is more pronounced in the curve for the copolymer (Fig. 3), and by comparison with the published work of Ambrose and Elliott, would indicate the presence of more of the folded structure.[2] Since it has been shown that the extended form predominates in short polypeptides,[7] which were used here, more striking shifts in the location of the peaks would be unexpected.

Discussion

Ambrose and Elliott[2] have pointed out that polypeptides of glycine are predominantly intermolecularly hydrogen bonded, and are soluble only in solvents which are capable of breaking these bonds.

It is at first surprising that the introduction of the large hydrophobic isobutyl side chain of leucine into the glycine polypeptide should increase its water solubility. However, this can be explained in part on the basis that the bulky side chain has reduced the amount of intermolecular hydrogen bonding. Consequently, fewer hydrogen bonds have to be broken in order to allow the molecules to pass into solution.

Bamford and associates,[7] on the basis of polarized infrared spectral data found that large side chains favored the folded form in polypeptides containing a single amino acid. It is of interest that similar results are obtained with polypeptides containing two amino acids and prepared in aqueous solutions. These findings, together with the observation that the polypeptides resulting from the copolymerizations are not only more soluble but also of higher average molecular weight than the polypeptides containing a single amino acid, gives support to the hypothesis[6] that the relative amounts of the folded and extended forms may be an important factor in determining the solubility of polypeptides. It would seem reasonable then to expect that the particular distribution of amino acids in the peptide chains of proteins would likewise affect the arrangement of the chains and the extent of intermolecular hydrogen bonding, and so influence the solubility behavior of the proteins.

MADISON 6, WISCONSIN

(11) We wish to thank Mr. Donald R. Johnson for making the infrared absorption measurements.

[CONTRIBUTION FROM THE DEPARTMENTS OF NEUROLOGY AND MICROBIOLOGY, COLLEGE OF PHYSICIANS AND SURGEONS, COLUMBIA UNIVERSITY, AND THE NEUROLOGICAL INSTITUTE, PRESBYTERIAN HOSPITAL]

Some Configurational Requirements and Dimensions of the Combining Site on an Antibody to a Naturally Occurring Antigen[1]

BY ELVIN A. KABAT

RECEIVED FEBRUARY 18, 1954

Oligosaccharides with 1 → 6 linkages inhibit the precipitation of human antidextran of 1 → 6 specificity and oligosaccharides with 1 → 4 specificity inhibit precipitation of antidextran with non 1 → 6 specificity by a dextran containing mainly 1 → 6 and 1 → 4 linkages. In each instance the homologous trisaccharide is much more effective than the homologous disaccharide. From the data on the inhibiting capacities of a variety of oligosaccharides of different structures the dimensions of the combining site of each type of antidextran are inferred to be complementary to an open chain of at least three α-D-glucopyranose units of homologous structure and probably to part or all of a fourth unit.

The classical studies of Landsteiner and co-workers[2] have established unequivocally the structural complementariness of the antibody combining group to the haptenic or determinant group of the antigen employed for immunization. These investigations as well as the more quantitative studies of subsequent workers[3,4] were carried out exclusively with artificial antigens prepared by introducing various determinant groups of known structure into the protein molecule. While the low molecular weight hapten introduced was capable of combining with the anti-hapten to inhibit precipitation of the the anti-hapten by a precipitating hapten or antigen, it was never possible to infer that any hapten used optimally satisfied the configuration requirements of the combining site on the antibody, since these may have been directed toward some unit larger than the group introduced into the antigen and may have involved not only the hapten but an indeterminate number of amino acids of the protein to which the hapten group had been attached. Indeed, Hooker and Boyd[5] showed that antibodies to a *p*-azophenylarsonate protein were inhibited

(1) This investigation was carried out under the William J. Matheson Commission and the Office of Naval Research (Contract #Nonr-266 (13)), Navy Department, Washington, D. C.

(2) K. Landsteiner, "The Specificity of Serological Reactions," Second Edition, Harvard University Press, Cambridge, Mass., 1945.

(3) L. Pauling, D. H. Campbell and D. Pressman, *Physiol. Rev.*, **23**, 203 (1943).

(4) D. H. Campbell and N. Bulman, "Progress in the Chemistry of Organic Natural Products," Vol. 9, L. Zechmeister, Editor, Vienna, 1952, p. 443.

(5) S. B. Hooker and W. C. Boyd, *J. Immunol.*, **25**, 61 (1933).