(42) G. W. Lehman and J. P. McTague, *J. Chem. Phys.*, **49**, 3170 (1968).
(43) A. Katchalsky and J. Gillis, *Recl. Trav. Chim. Pays-Bas*, **68**, 879 (1949).
(44) A. Arnold and J. Overbeek, *Recl. Trav. Chim. Pays-Bas*, **69**, 192 (1950).
(45) A. Yaron, N. Tal, and A. Berger, *Biopolymers*, **11**, 2461 (1972).
(46) R. G. Bates, "Determination of pH", Wiley, New York, N.Y., 1964, p 189.
(47) H. O. Spivly and T. Shedlovsky, *J. Phys. Chem.*, **71**, 2171 (1967).
(48) H. O. Spivly and T. Shedlovsky, *J. Phys. Chem.*, **71**, 2165 (1967).
(49) V. E. Bychkova and O. B. Ptitsyn, *Biopolymers*, **10**, 2161 (1970).
(50) A. W. Burgess, P. K. Ponnuswamy, and H. A. Scheraga, *Isr. J. Chem.*, **12**, 239 (1974).
(51) P. Chou and G. D. Fasman, *Biochemistry*, **13**, 211 (1974).
(52) P. Doty, A. Wada, J. T. Yang, and E. R. Blout, *J. Polym. Sci.*, **23**, 851 (1957).
(53) S. Lewin, *J. Theor. Biol.*, **36**, 1 (1972).
(54) R. M. Ispiryan, L. I. Brelbnkii, and Ya. L. Goldfarb, *Izv. Akad. Nauk Kaz. SSR, Ser. Khim.*, **11**, 2513 (1967).
(55) J. R. Piper, L. R. Stringfellow, Jr., and T. P. Johnson, *J. Heterocycl. Chem.*, **4**, 298 (1967).
(56) "CRC Handbook of Chemistry and Physics", 54th ed, Chemical Rubber Publishing Co., Cleveland, Ohio, 1973, p C-221.
(57) Reference 23, p 29.
(58) E. Schroder and E. Klieger, *Justus Liebigs Ann. Chem.*, **673**, 196 (1964).
(59) M. Rinaudo, A. Domard, and G. Spach, *Biopolymers*, **12**, 2199 (1973).

The Effect of Neighboring Charges on the Helix Forming Ability of Charged Amino Acids in Proteins[1]

Frederick R. Maxfield[2] and Harold A. Scheraga*

Department of Chemistry, Cornell University, Ithaca, New York 14853.
Received February 10, 1975

ABSTRACT: It has been found that the fraction of glutamic acid residues which are helical in proteins is larger than might be expected from the experimentally determined value of the helical stability constants of glutamic acid. In order to understand this difference, the effect of neighboring charged side chains on the glutamic acid residues in proteins of known structure is examined. It is found that a positively charged side chain four residues away from a glutamic acid greatly enhances its probability to be helical. Similar results are obtained for aspartic acid, lysine, arginine, and histidine.

The helix–coil stability constants for charged poly(L-glutamic acid) in 0.1 N KCl have been determined[3] experimentally using the host–guest technique[4-11] under conditions where the long-range electrostatic interactions do not contribute significantly to the helix–coil transition. These results indicated that an isolated charged glutamic acid residue at 20° is essentially indifferent as far as its helix forming or breaking tendency is concerned. On the other hand, recent surveys of several proteins whose crystal structures have been determined indicate that glutamic acid has the highest probability to be helical of all the amino acids.[12,13] In comparing the helical stability in random copolymers with the probability that a residue will be helical in a protein, it is important to remember that the former seems to be determined by short-range interactions,[4,5,14] while the latter will also include the effects of specific medium- and long-range interactions.[12,15] In this paper, it is shown by a statistical analysis of the conformations of amino acid residues in proteins of known structure that a specific medium-range interaction (viz., the presence of a positively charged side chain four residues away from a glutamic acid residue in a protein) greatly enhances the likelihood that glutamic acid will be helical. It is also shown that this type of interaction (i.e., with oppositely charged side chains) influences the conformational preference of aspartic acid, lysine, arginine, and histidine.

I. Methods

The proteins used in this investigation are the 14 listed in Table V of ref 12 plus bovine pancreatic trypsin inhibitor[16] and trypsin.[17] The definition of an α helix is that used by Burgess et al.,[12] and various frequencies of occurrence (as indicated below) were recorded.

The statistical significance of any difference in the likelihood for a residue to be helical in the presence or absence of a specific interaction was determined using Fisher's "goodness of fit" criterion, based on an exact treatment of 2×2 contingency tables.[18] The four entries in each contingency table were the number of times, $n_{\alpha\beta}$, that a given amino acid residue was found to occur in the data set with conditions specified by the subscripts. When $\alpha = 1$, say, the residue being studied is helical, and when $\alpha = 2$, it is not helical. When $\beta = 1$, the type of medium-range interaction being studied is present, and it is absent when $\beta = 2$. For example, if the ith residue is glutamic acid, $\beta = 1$ may mean that the $(i + 4)$th residue is positively charged, whereas $\beta = 2$ indicates that it is not positively charged. If the interaction being studied does not affect the likelihood for the amino acid under consideration to be helical, one would expect n_{11}/n_{21} to be roughly the same as n_{12}/n_{22}. Fisher's method measures the probability, P, that deviations of n_{11}/n_{21} from n_{12}/n_{22} as large or larger than those actually observed in the data could occur by chance; e.g., if P is 0.01, then the probability is 1 in 100 that such a difference could occur by chance.

In order to evaluate P, it is necessary to consider the contingency table further. With the margins of the table fixed [i.e., with $(n_{11} + n_{12})$, $(n_{11} + n_{21})$, $(n_{12} + n_{22})$, and $(n_{21} + n_{22})$ fixed], one can calculate the values expected for each entry in the table if the conditions specified for the rows (i.e., values of β) do not affect the distribution between the columns (i.e., values of α). We define an expected frequency $e_{\alpha\beta}$ as

$$e_{\alpha\beta} = (n_{\alpha 1} + n_{\alpha 2})(n_{1\beta} + n_{2\beta})/(n_{11} + n_{12} + n_{21} + n_{22}) \quad (1)$$

Further, it is convenient to define a quantity, Q, by the equations

$$Q = \sum_{h=0}^{m} |(n_{11} - h)!(n_{12} + h)!(n_{21} + h)!(n_{22} - h)!|^{-1} \quad (2a)$$

for $n_{11} \leq e_{11}$

Table I
Comparison of the Zimm-Bragg Parameter[19] s with the Fraction of Times That an Amino Acid Residue is Found to be Helical in Proteins

Amino acid	s_{20}[a]	F_α[b]
Leucine	1.14[9]	0.33
Phenylalanine	1.09[10]	0.28
Alanine	1.07[7]	0.39
Glutamic acid	0.97[3]	0.47
Valine	0.93[11]	0.26
Serine	0.78[8]	0.13
Glycine	0.59[6]	0.09

[a] Determined by the host-guest technique as described in the references cited. The solvent was water in each case except for glutamic acid where the polymers were dissolved in 0.1 N KCl adjusted to pH 8 with KOH. [b] F_α is the number of times that an amino acid residue was found to be part of a right-handed α helix divided by the total number of times that the amino acid residue occurred in all of the proteins studied.

and

$$Q = \sum_{h=0}^{l} |(n_{11}+h)!(n_{12}-h)!(n_{21}-h)!(n_{22}+h)!|^{-1} \quad (2b)$$

$$\text{for } n_{11} > e_{11}$$

In eq 2a and 2b, m is the smaller of the two numbers n_{11} and n_{22}, and l is the smaller of n_{12} and n_{21}. Then,[18]

$$P = Q(n_{11}+n_{12})!(n_{21}+n_{22})!(n_{11}+n_{21})!\,(n_{12}+n_{22})!/(n_{11}+n_{12}+n_{21}+n_{22})! \quad (3)$$

II. Results and Discussion

Table I lists the values of the Zimm–Bragg parameter[19] s at 20° determined by the host–guest technique and the values of F_α, the number of times that an amino acid residue was found to be part of a right-handed α helix divided by the total number of occurrences of that amino acid residue in all of the proteins studied. It is interesting to note that, for six of the naturally occurring amino acids studied previously by the host–guest technique, there is a reasonable monotonic correlation between the Zimm–Bragg[19] parameter s and F_α (i.e., those amino acids with high values of s have relatively high probabilities of being helical in proteins). This correlation reflects the importance of short-range interactions[14] in determining the helical stability of these amino acids in proteins.

Glutamic acid appears to constitute an exception to this correlation. The fact that glutamic acid has a high value of F_α and an intermediate value of s may indicate that medium- and long-range interactions influence the conformation of glutamic acid residues in proteins. In order to investigate this possible effect of medium-range electrostatic interactions, the probabilities for a glutamic acid residue to be helical when certain of its neighbors are charged are compared with the probabilities to be helical when the neighbors are not charged. The results are summarized in Tables II and III. In Table II the effect of a lysine, arginine, or histidine j residues away from a glutamic acid residue at position i in the chain is tabulated. For example, it was found that 23 glutamic acid residues had at least one nearest neighbor ($j = 1$) which was lysine, arginine, or histidine, and in 13 of those cases the glutamic acid was helical. Of the remaining 82 glutamic acid residues, 36 were helical. For the case of $j = 1$, $n_{11} = 13$, $n_{12} = 36$, $n_{21} = 10$, $n_{22} = 46$, $n_{11} + n_{12} = 49$, $n_{21} + n_{22} = 56$, $n_{11} + n_{21} = 23$, $n_{12} + n_{22} = 82$, and $n_{11} + n_{12} + n_{21} + n_{22} = 105$. The quantity e_{11} is 10.7; hence, since $n_{11} > e_{11}$, eq 2b is used to obtain Q (with $l = 10$) as 4.10×10^{-116}. Then, from eq 3,

$$P = 4.10 \times 10^{-116} \times 49!56!23!82!/105! \simeq 0.20 \quad (4)$$

Table II
Effect of a Lysine, Arginine, or Histidine j Residues Away from a Glutamic Acid at Position i in Proteins

j	Fraction of Glu that is helical if $i + j$ or $i - j$ is Lys, Arg, or His[a]	Fraction of Glu that is helical if $i + j$ and $i - j$ are not Lys, Arg, or His[a]	P[b]
1	13/23 (0.57)	36/82 (0.44)	0.20
2	16/29 (0.55)	33/76 (0.43)	0.20
3	16/29 (0.55)	33/76 (0.43)	0.20
4	24/37 (0.65)	25/68 (0.37)	0.005
5	9/25 (0.36)	40/80 (0.50)	0.16

[a] Based on a survey of 16 proteins as described in the text. [b] P is a measure of the probability that the observed distribution of helical residues between the two columns could occur by chance. A small value indicates that it is unlikely that the deviation from proportionality is due to randomness.[18]

The values indicated for the test of significance are the probability, P, that a departure from exact proportionality between the two columns as large or larger than that shown (i.e., 0.57/0.44 for $j = 1$) could occur by chance. Table III is similar to Table II, except that the effect of other negatively charged side chains near to a glutamic acid residue is shown.

It can be seen that the only significant deviation from proportionality in the two tables (Tables II and III) occurs when $i + 4$ or $i - 4$ is lysine, arginine, or histidine. A glutamic acid residue is significantly more likely to be helical when at least one of the side chains four residues to either side of it has a positive charge. For the data used in our analysis, nearby negative charges do not influence its conformation significantly (Table III). On a molecular level the effect of a positive side chain four residues from a glutamic acid residue is reasonable since an α helix brings the side chains of the $i \pm 4$ residue close to the side chain of residue i and allows for a favorable electrostatic interaction. This type of interaction would explain the fact that glutamic acid has a large value of F_α and an intermediate value of the Zimm–Bragg parameter s since 24 out of the 49 glutamic acid residues which are helical in these proteins have a lysine, arginine, or histidine at $i + 4$ or $i - 4$, and s is not influenced by long-range electrostatic interactions in the copolymers studied.[3]

Similar analyses were carried out for aspartic acid, lysine, arginine, and histidine, and the results are summarized in Table IV for the case when $i \pm 4$ is of the opposite charge from the residue at position i. In general, the results for these amino acids were similar to those for glutamic acid in that significant departures from proportionality were observed only when $i + 4$ or $i - 4$ carries the opposite charge from the ith residue. Exceptions occurred for lysine which was also more likely to be helical when residues $i \pm 2$ ($P = 0.007$) or $i \pm 3$ ($P = 0.02$) were negatively charged and for aspartic acid which was more likely to be helical when residue $i \pm 3$ ($P = 0.04$) was positively charged. Of the five amino acids examined, only lysine showed any significant difference between the effects on the left from those on the right, i.e., between the $i + 4$ and the $i - 4$ positions. In the 22 cases where $i - 4$ was negatively charged (i.e., glutamic acid or aspartic acid), 14 of the lysine residues were helical; while in the 26 cases where $i + 4$ was negatively charged,

Table III
Effect of a Glutamic Acid or Aspartic Acid *j* Residues Away from a Glutamic Acid at Position *i* in Proteins

j	Fraction of Glu that is helical if $i + j$ or $i - j$ is Glu or Asp[a]	Fraction of Glu that is helical if $i + j$ and $i - j$ are not Glu or Asp[a]	P[b]
1	13/24 (0.54)	36/81 (0.44)	0.27
2	16/29 (0.55)	33/76 (0.43)	0.20
3	11/25 (0.44)	38/80 (0.48)	0.47
4	5/17 (0.29)	44/88 (0.50)	0.10
5	10/18 (0.56)	39/87 (0.45)	0.28

[a] Same as in Table II. [b] Same as in Table II.

Table IV
Effect of an Amino Acid Four Residues Away from Position *i* Which Carries the Opposite Charge from That at Position *i*[a]

Residue *i*	Fraction helical if $i + 4$ or $i - 4$ carries the opposite charge[b]	Fraction helical if neither $i + 4$ nor $i - 4$ carries the opposite charge[b]	P[c]
Asp	12/28 (0.43)	17/96 (0.18)	0.008
Lys	20/44 (0.45)	33/137 (0.24)	0.007
Arg	8/13 (0.62)	9/59 (0.15)	0.001
His	6/12 (0.50)	10/45 (0.22)	0.07

[a] For this analysis histidine was considered to carry a positive charge. [b] Based on a survey of 16 proteins as described in the text. [c] See footnote *b* in Table II.

only eight were helical. Thus, a negative charge at the $(i - 4)$th residue is more influential than one at the $(i + 4)$th residue in causing a lysine to be helical. The test of significance gave a value of 0.02 for this difference in behavior at the $(i - 4)$th and $(i + 4)$th positions.

This analysis indicates that the likelihood for a charged residue in a protein to be helical is strongly influenced by the possibility of a favorable interaction with side chains of the opposite charge, particularly when the opposite charges are four residues away on either side. Such medium-range interactions were implicit in the nonamer calculations of Ponnuswamy et al.[15] It may prove useful to incorporate this type of information in algorithms used to predict the backbone structure of proteins from the amino acid sequence.[12]

Acknowledgment. We are grateful to Dr. A. W. Burgess for helpful discussions.

References and Notes

(1) This work was supported by research grants from the National Institute of Arthritis and Metabolic Diseases, U.S. Public Health Service (AM-08465), and from the National Science Foundation (BMS-71-00872-A04).
(2) NSF Predoctoral Trainee, 1971–1974; NIH Predoctoral Trainee, 1974–1975.
(3) F. R. Maxfield, J. E. Alter, G. T. Taylor, and H. A. Scheraga, *Macromolecules*, preceding paper in this issue.
(4) P. H. Von Dreele, D. Poland, and H. A. Scheraga, *Macromolecules*, **4**, 396 (1971).
(5) P. H. Von Dreele, N. Lotan, V. S. Ananthanarayanan, R. H. Andreatta, D. Poland, and H. A. Scheraga, *Macromolecules*, **4**, 408 (1971).
(6) V. S. Ananthanarayanan, R. H. Andreatta, D. Poland, and H. A. Scheraga, *Macromolecules*, **4**, 417 (1971).
(7) K. E. B. Platzer, V. S. Ananthanarayanan, R. H. Andreatta, and H. A. Scheraga, *Macromolecules*, **5**, 177 (1972).
(8) L. J. Hughes, R. H. Andreatta, and H. A. Scheraga, *Macromolecules*, **5**, 187 (1972).
(9) J. E. Alter, G. T. Taylor, and H. A. Scheraga, *Macromolecules*, **5**, 729 (1972).
(10) H. E. Van Wart, G. T. Taylor, and H. A. Scheraga, *Macromolecules*, **6**, 266 (1973).
(11) J. E. Alter, R. H. Andreatta, G. T. Taylor, and H. A. Scheraga, *Macromolecules*, **6**, 564 (1973).
(12) A. W. Burgess, P. K. Ponnuswamy, and H. A. Scheraga, *Isr. J. Chem.*, **12**, 239 (1974).
(13) P. Chou and G. D. Fasman, *Biochemistry*, **13**, 211 (1974).
(14) H. A. Scheraga, *Pure Appl. Chem.*, **36**, 1 (1973).
(15) P. K. Ponnuswamy, P. K. Warme, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.*, **70**, 830 (1973).
(16) R. Huber and W. Steigemann, private communication.
(17) (a) R. M. Stroud, L. M. Kay, and R. E. Dickerson, *J. Mol. Biol.*, **83**, 185 (1974); (b) the coordinates for trypsin were supplied in a private communication from R. M. Stroud.
(18) R. A. Fisher, "Statistical Methods for Research Workers", Oliver and Boyd, London, 1958, p 96.
(19) B. H. Zimm and J. K. Bragg, *J. Chem. Phys.*, **31**, 526 (1959).