Analysis of the Code Relating Sequence to Conformation in Globular Proteins

AN INFORMATIONAL ANALYSIS OF THE ROLE OF THE RESIDUE IN DETERMINING THE CONFORMATION OF ITS NEIGHBOURS IN THE PRIMARY SEQUENCE

By BARRY ROBSON and ROGER H. PAIN
*Department of Biochemistry, University of Newcastle upon Tyne,
Newcastle upon Tyne NE1 7RU, U.K.*

(Received 24 September 1973)

1. The effect exerted by a residue on the conformation of neighbouring residues was analysed by using data from nine globular proteins of known sequence and conformation. 2. An information measure was used which estimated the role of a residue in influencing neighbouring conformations and also its tendency to influence the lengths of runs of residues in that conformation. This measure was estimated for each residue in all conformations defined by domains on the ϕ, ψ diagram. 3. Plots of the information measure yielded an intercept, which was a measure of intra-residue information for a residue. The slope was a measure of the statistical co-operativity or tendency of the residue to influence the occurrence of its neighbours in runs of a particular conformation. Both parameters are a function of the residue type. Statistical co-operativity is found in the α_1 -helical (H_1) and β -pleated-sheet (P_2) conformations and, to a lesser extent, in their distorted variants H_2 and P_1 . 4. The directional nature of these influences for H_1 and P_2 conformations is illustrated by plots of the information measure against the distance m from the residue, for $m = -10$ to $+10$. 5. The results for statistical co-operativity are discussed in relation to theories of helix-coil and pleated-sheet-coil transitions. The value of the information-theory-derived parameters in obtaining s parameters for the Zimm & Bragg (1959) equations is illustrated. 6. Directional effects are discussed with particular relation to mechanisms of the termination of helices and the involvement of the α_{II} conformation and also to discontinuities in pleated-sheet conformations.

A typical residue in a globular protein molecule has an important but not decisive role in determining its own backbone conformation (Guzzo, 1965; Prothero, 1966; Havsteen, 1966; Ptitsyn, 1969; Kotelchuck & Scheraga, 1968, 1969; Pain & Robson, 1970; Robson & Pain, 1971). This role has been described as involving intra-residue information which has been measured by analysis of proteins of known sequence and conformation (Robson & Pain, 1974a). Since the information that determines the native state under biological conditions is believed to be contained entirely in the primary sequence (Anfinsen, 1967; Tanford, 1968), it follows that the other residues in the sequence contribute to that typical residue the remaining information required to define its conformation correctly and unambiguously. Since this further information involves at least two residues, namely the residue whose side chain is the information source and the residue whose conformation is the information 'sink', it is conveniently referred to as inter-residue information. Information-theory analysis enables the independent contributions of statistically interdependent events to be measured; therefore it is possible to determine the

contribution of inter-residue information by analysis of proteins of known sequence and conformation.

Similarly, information-theory analysis makes it possible to break down the inter-residue information into still more specific types of contribution. In particular, the residue whose conformation is the information 'sink' may have a role in modifying the nature and magnitude of the information contributed by neighbours, but there may be a fraction of that contribution which is not so modified. For example, the exact effect that the presence of the proline residue has on the conformation of the residues on its *N*-terminal side will depend in part on the types of amino acid residue found on that side, but proline will also have some influence on the tendency of those residues to occur in the right-hand α -helix which is quite independent of the nature of those residues (Robson & Pain, 1972). It is the latter independent or 'unconditional' (as opposed to the dependent or 'conditional') information that is concerned with the important phenomenon of co-operativity and related effects. It controls the tendency of some conformations of certain residues to occur in runs rather than singly and the tendency

of some residues to allow the spread of certain kinds of run to develop in one direction but not in another. Such effects, which are certainly important in the formation of the right-hand α -helix in proteins (Ptitsyn, 1969; Robson & Pain, 1972), are investigated in the present paper and extended for the first time to other conformations.

Theory

We are interested in measuring the information carried by a residue of type R at position j concerning the conformation S at position $j+m$. Call this information $I(S_{j+m}; R_j)$ and note the special case $I(S_j; R_j)$, which is the definition of intra-residue information (Robson & Pain, 1974a). The conformation S_{j+m} may be described by a two-valued scalar quantity which is X if the residue backbone angles belong to a specified domain (see Fig. 1), and $\bar{X}$ if they do not, and the above information measure becomes $I(S_{j+m} = X; \bar{X}; R_j)$.

Each residue with a particular conformation X will, in general, belong to a run of residues of similar conformation, though this run may in the limit be only one residue long. Thus, conformation S_{j+m} will be an element of some string of $n \geq 1$ residue conformations:

$$S_{j+m} = S_i \dots S_{j+m} \dots S_{i+n-1} \quad (1)$$

all with the same conformation X . The string of

Fig. 1. Definition of conformation domains

Domains defined on the basis of the grouping of information maxima derived from the natural distribution of residues (Robson & Pain, 1974a).

conformations S_{j+m}^n may, by analogy with S_{j+m} , be described by a vector of two-valued elements which is $\{X\}$ if all the elements are X , and $\{\bar{X}\}$ if not all the elements are X . $\{\bar{X}\}$ is to be distinguished from $\{X\}$, which is a run of conformations of which none is X . By analogy with $I(S_{j+m} = X; \bar{X}; R_j)$ we may write $I(S_{j+m}^n = \{X\}; \{\bar{X}\}; R_j)$. This is the information that a residue R_j gives about the occurrence of a residue at position $j+m$ in a run of n or more residues in the conformation X as opposed to the case where one or more residues in the run is $\bar{X}$. Note the special case $n=1$ when $j+m$ and $i+n-1$ are the same position and when $I(S_{j+m}^{n-1} = \{X\}; \{\bar{X}\}; R_j) = I(S_j = X; \bar{X}; R_j)$.

It is implicit that the string of conformations S_{j+m}^n constitutes a substring of $S_{j+m}^{n'}$ if $n < n'$; that is, a conformation which belongs to a string of X types of length greater than n must also belong to a string of X types of length n . In other words, S_{j+m}^n does not necessarily imply that a string is bounded at each end by a conformation $\bar{X}$ rather than X . The parameter n therefore refers to the minimum length of any run and the above definitions are generalizations of those already in use, either implicitly or explicitly. For example, a residue is usually defined as α -helical only if it belongs to a run of at least $n=4$ residues in the α -helical conformation (Phillips, 1967).

It should be noticed that when applied to the definition of β -pleated-sheet conformations, no account is taken of the hydrogen-bonding arrangement and a residue can be defined therefore as belonging to a string of residues in the β -pleated-sheet conformation, even if no clear pleated-sheet or ribbon structure is formed.

Methods

Information for two possible outcomes $\{X\}$ and $\{\bar{X}\}$, can be re-expressed in terms of 'star' functions $\star(\quad)$ which are the terms obtained directly by statistical analysis of proteins of known sequence and conformation. A simple relation exists between I and $\star$ functions:

$$\begin{aligned} \star(S_{j+m}^n = \{X\}; \{\bar{X}\}; R_j) \\ = I(S_{j+m}^n = \{X\}; \{\bar{X}\}; R_j) \\ + I(S_{j+m}^n = \{X\}; \{X\}) \end{aligned} \quad (2)$$

This expresses the preference that a residue at position $j+m$ has for being in state $\{X\}$ when a residue of type R occurs at j . This preference is the complete information which is the sum of $I(S_{j+m}^n = \{X\}; \{\bar{X}\}; R_j)$, the information contained in R_j as to whether or not residue j will be in state $\{X\}$, and $I(S_{j+m}^n = \{X\}; \{X\})$, the information which would be available about the state of the residue at position $j+m$ even before residue R_j is taken into account. The function:

$$I(S_{j+m}^n = \{X\}; \{\bar{X}\}) = \star(S_{j+m}^n = \{X\}; \{\bar{X}\}) \quad (3)$$

is non-zero because, for example, we know that any residue is less likely to occur in a α -helix simply because there are more residues outside than inside the α -helical regions of proteins of known sequence and conformation.

In this work the information $I(\$_{j+m}^n = \{X\}; \{\bar{X}\}; R_j)$ is estimated as $E[I(\$_{j+m}^n = \{X\}; \{\bar{X}\}; R_j) | D]$ for different values of m and n , where D is the data pool from which the conformational preferences are drawn. The evaluation of this estimate is analogous to that of $E[I(S_j = X; \bar{X}; R_j) | D]$ described by Robson & Pain (1974a), and values were obtained by analysis of nine proteins of known sequence and conformation, as listed in Robson & Pain (1974a).

$\star(\$_{j+m}^n = \{X\}; \{\bar{X}\})$ is similarly estimated as $E[\star(\$_{j+m}^n = \{X\}; \{\bar{X}\}) | D]$.

Because of the large numbers of combinations of X , m and n , the following approach was adopted. First, the value of n was varied with $m = 0$, for all the different conformations X . The negative value of slope, $-d\{E[I(\$_{j+m}^n = \{X\}; \{\bar{X}\}; R_j) | D]\}/dn$, at each point is the amount of information associated with adding one more residue conformation to a string of residue conformations of which one is the conformation of R_j . This information will be small when a conformation X has a greater than average tendency to occur in runs, since the extension of the run is not then a rare event. Such a conformation X may be described as statistically co-operative. In this way, it is possible to spot those conformations X which possess statistical co-operativity and, incidentally, those amino acid residues R_j which tend to occur in shorter or longer runs.

Secondly, by using $m = 0$ and those conformations X which exhibited statistical co-operativity, the value of n was found which, for each conformation, gave intra-residue information measures leading to optimal predictions of runs. Consider, for example, the prediction of right-hand α -helix. When $n = 1$, many residues are included that possess helical ϕ - ψ angles, but which belong to the coil region of the sample. Certain amino acid residues which do not like to occur within a run of α -helix might then be given too high a helix-forming information measure which would decrease the accuracy of prediction of helical regions. When $n = 20$, many residues which belong to the α -helical regions are excluded and the sample becomes small and unreliable. Somewhere between will be an optimal value of n .

The predictions were carried out as follows. For $m = 0$, for each value of n , for each type of amino acid, and for a given conformation X , a value of $E[I(\$_{j+m}^n = \{X\}; \{\bar{X}\}; R_j) | D]$ was generated. This term represents the information given by the residue R_j as to its occurrence in a string of n residues having the conformation X . By using this intra-residue

information and the same value of n , predictions of each residue as X or $\bar{X}$ were made in the nine proteins of known sequence and conformation. A residue was predicted to be of conformation X if the residue belonged to a run of n residues with a total information content greater than an empirically determined decision constant, this procedure being described in detail by Robson (1974).

Finally, having identified the conformations X which exhibit structural co-operativity, and the optimal value of n associated with each of those conformations, the estimate $E[I(\$_{j+m}^n = \{X\}; \{\bar{X}\}; R_j) | D]$ was evaluated for all values of m from -10 to $+10$.

Results

The values of $E[I(\$_{j+m}^n = \{X\}; \{\bar{X}\}; R_j) | D]$ were plotted against n . Typical examples are shown in Fig. 2 for sub-domain H₁ (right-hand α -helix). The slopes are negative because less residues are defined as helical when n is increased. n is the minimum number of residues in a helical conformation which must occur in a run for that run to be termed a helix. In each case, a plot is produced which is close to linearity for values of n greater than 4. This reflects the fact that the information associated with pulling one more residue into the same conformation as the string is generally independent of the length of long strings. The plots continue to be linear up to a value of n above which no strings are found. At this point there is a discontinuity and the plot becomes horizontal. The statistical co-operativity can thus be described by the negative value of the gradient. The results for all conformations and all residues are shown in Table 1.

A small value of the gradient implies that only a small value of information is required to add a

Fig. 2. Plots of $I_c = E[\star(\$_{j+m}^n = \{X\}; \{\bar{X}\}); R_j) | D]$ versus n for alanine (○—○), serine (○—○) and proline (●).

For details see the text.

Table 1. Parameters describing the plots of $E[\{X\}:\{\bar{X}\};R_j|D]$ versus length of residue string, n

For details see the text.

Residue	H ₁		H ₂		P ₁		P ₂	
	Slope (nats)	Intercept (nats)	Slope (nats)	Intercept (nats)	Slope (nats)	Intercept (nats)	Slope (nats)	Intercept (nats)
Gly	-0.08	-2.0	-1.50	-2.4	-1.6	-2.0	-0.70	-1.7
Ala	-0.14	-0.1	-0.58	-1.7	-0.90	-1.7	-0.37	-1.2
Val	-0.14	-0.7	-1.10	-1.9	-0.88	-1.6	-0.78	-0.2
Leu	-0.08	-0.7	-1.40	-1.5	-0.56	-1.3	-0.7	-0.75
Ile	-0.15	-0.35	-1.61	-1.6	-0.96	-1.5	-0.74	-0.25
Ser	-0.21	-1.0	-1.07	-1.4	-1.43	-1.0	-0.65	-1.1
Thr	-0.18	-1.1	-1.3	-1.2	-1.30	-1.3	-0.65	-1.1
Asp	-0.13	-0.6	<-2	-1.0	<-2	-2.4	-0.76	-0.8
Glu	-0.14	+0.2	-0.75	-1.35	-0.96	-1.9	-1.0	-1.6
Asn	-0.23	-1.13	-1.17	-1.3	-1.55	-1.8	-0.60	-1.9
Gln	-0.26	-1.0	-1.46	-1.6	-1.08	-1.3	-1.0	-0.4
Lys	-0.12	-0.3	-0.85	-1.8	-1.75	-1.7	-0.45	-1.6
His	-0.10	-0.65	-1.46	-1.3	-1.30	-1.6	-0.95	-2.3
Arg	-0.16	-0.7	-1.07	-1.4	-1.20	-1.0	-0.82	-1.3
Phe	-0.19	-0.6	-1.20	-2.0	-1.40	-1.6	-0.72	-0.8
Tyr	-0.14	-1.6	-1.00	-1.4	<-2	-1.6	-0.66	-0.6
Trp	-0.16	-0.8	-0.80	-0.8	<-2	-2.0	-0.58	-1.1
Cys	-0.24	-1.1	<-2	-0.7	-1.25	-1.8	-0.78	-0.4
Met	-0.21	-0.2	-1.1	-2.5	-0.9	-1.6	-0.80	-2.5
Pro	-0.32	-1.1	-1.5	-1.2	-1.1	0.0	<-2	-4.4

residue conformation X to a string of residue conformations X ; this is characteristic of a high degree of co-operativity possessed by R_j in conformation X . This co-operativity, as measured by the gradient, is not a function of the type of residue added: the latter represents a statistical set of residues. For example, when a residue which has a preference for occurring at the boundary of a helical run occurs at the boundary, the helix will tend to be shorter, and a residue of strong helix-forming power must be brought up if the helix is to be extended.

Only conformations X belonging to sub-domains H_1 and P_2 (Robson & Pain, 1974a) were strongly co-operative. These sub-domains correspond to right-hand α -helix and pleated sheet respectively. Sub-domains H_2 and P_1 showed much lower statistical co-operativity. These sub-domains correspond to distorted right-hand α -helix (including 3_{10} , α_{II} , and some reverse turns; see Robson & Pain, 1974b) and distorted pleated sheet respectively. All the other sub-domains produced almost vertical plots which entered the discontinuity and became horizontal at very low values of n . This is also true of L_1 , the sub-domain corresponding to the left-hand α -helix. In this case, the absence of left-hand α -helix in the proteins of known conformation is not simply due to the infrequency of residues in the L_1 conformation (as revealed in the intercept on the ordinate), but also to the physical instability of runs of this

conformation (as revealed by the steepness of the gradient).

Fig. 2 and Table 1 reveal that the nature of residue R_j has an effect on the gradient of the plots. This reflects the fact that residues which act, for example, as helix-breakers and occur at helix boundaries will tend to occur in shorter helices.

The value of n was optimized for the strongly co-operative conformations H_1 and P_2 . The optimal value of n was 4 in both cases and the number of residues correct was 76% and 78% respectively. By using this value of n , values of $E[I(\{X\}_{j+m}^4 = \{X\}:\{\bar{X}\};R_j)|D]$ were generated for conformations H_1 and P_2 and for $-10 \leq m \leq +10$. These values are listed in Table 2, and represent the information that a residue at a distance m from residue R_j will be one of a run of at least four residues in conformation X .

It is possible to classify directional effects, such as are represented by Table 2, into five main groups (cf. Robson & Pain, 1972).

Group I consists of information estimates of $I(\{X\}_{j+m}^4 = \{X\}:\{\bar{X}\};R_j)$ which form a maximum at $m = 0$.

Group II consists of estimates which form a minimum at $m = 0$.

Group III displays a maximum for a value of m greater than 0 and a minimum for a value of m less than 0.

Table 2. Values of the information measure $E[\star(S_{i+m}^4 = \{X\}; \{X\}; R_i)D]$ evaluated as a function of residue spacing, m , for the two statistically co-operative conformations, H_1 and P_2

The measures are evaluated in decinats, for each conformation at the optimal value of $n = 4$. {To obtain $E[I(S_{i+m}^4 = \{X\}; \{X\}; R_i)D]$ as plotted in Figs. 3-8, add 1.2nats to the values for conformation H_1 and 2.5nats to the values for P_2 (see the text).}

Conformation H_1		Residue	m	...	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
		Gly			-17	-17	-14	-15	-15	-17	-16	-19	-20	-22	-22	-20	-20	-18	-15	-13	-14	-15	-14	-12	-15
		Ala			-12	-10	-10	-9	-8	-7	-7	-8	-5	-5	-5	-6	-5	-6	-9	-10	-10	-10	-11	-11	-11
		Val			-14	-15	-16	-15	-15	-14	-14	-12	-15	-14	-12	-13	-11	-12	-11	-13	-13	-17	-16	-17	-15
		Leu			-10	-9	-10	-10	-11	-9	-9	-8	-9	-10	-11	-11	-10	-10	-11	-8	-8	-10	-10	-13	-12
		Ile			-14	-10	-13	-10	-9	-11	-12	-12	-12	-11	-9	-7	-9	-7	-8	-8	-12	-10	-12	-13	-11
		Ser			-18	-17	-17	-17	-17	-18	-19	-19	-19	-19	-20	-17	-17	-15	-13	-15	-15	-16	-15	-15	-16
		Thr			-16	-16	-17	-15	-17	-18	-19	-19	-19	-17	-17	-16	-17	-16	-15	-13	-14	-14	-13	-13	-14
		Asp			-13	-13	-11	-11	-10	-9	-11	-14	-15	-12	-11	-9	-8	-7	-8	-11	-11	-11	-13	-12	-12
		Glu			-3	-5	-3	-8	-9	-10	-7	-7	-2	-3	-1	-2	-3	-2	-1	-3	-3	0	-2	-3	-4
		Asn			-15	-16	-17	-16	-17	-14	-12	-13	-12	-14	-16	-16	-18	-17	-19	-20	-19	-19	-23	-21	-25
		Gln			-20	-20	-20	-19	-21	-18	-19	-18	-16	-16	-18	-16	-18	-18	-21	-18	-17	-17	-13	-13	-14
		Lys			-5	-5	-3	-3	-4	-3	-4	-3	-4	-5	-6	-8	-11	-11	-11	-11	-9	-10	-10	-11	-9
		His			-4	-6	-6	-5	-4	-5	-5	-5	-6	-7	-11	-11	-9	-11	-9	-9	-7	-9	-7	-4	-4
		Arg			-20	-20	-17	-17	-17	-17	-15	-13	-13	-12	-12	-12	-20	-20	-18	-18	-17	-12	-14	-13	-16
		Phe			-8	-8	-7	-9	-6	-9	-8	-10	-8	-10	-10	-12	-8	-12	-10	-15	-13	-11	-10	-8	-9
		Tyr			-13	-15	-15	-18	-16	-18	-19	-21	-21	-24	-20	-18	-15	-14	-18	-16	-13	-13	-11	-9	-9
		Trp			-23	-20	-17	-14	-13	-17	-20	-15	-11	-11	-11	-11	-5	-9	-9	-11	-17	-20	-23	-20	-20
		Cys			-25	-25	-30	-30	-30	-25	-25	-19	-19	-19	-17	-22	-22	-25	-25	-30	-30	-30	-40	-40	-40
		Met			-14	-14	-14	-7	-9	-5	-3	-3	-3	-5	-5	-3	-5	-12	-9	-15	-15	-15	-14	-18	-18
		Pro			-17	-19	-21	-24	-27	-24	-27	-27	-24	-28	-24	-19	-17	-17	-12	-11	-12	-11	-12	-12	-10
Conformation P_2		Residue	m	...	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
		Gly			-29	-31	-27	-31	-22	-21	-18	-19	-22	-29	-43	-53	-53	-31	-27	-26	-24	-24	-24	-21	-21
		Ala			-29	-29	-31	-37	-34	-27	-27	-25	-26	-25	-23	-25	-26	-34	-29	-24	-24	-23	-26	-31	-33
		Val			-23	-24	-22	-21	-24	-25	-35	-32	-25	-18	-17	-19	-20	-25	-27	-32	-32	-27	-32	-29	-35
		Leu			-26	-28	-30	-30	-30	-23	-23	-20	-19	-21	-21	-26	-30	-28	-30	-28	-28	-30	-28	-28	-35
		Ile			-32	-28	-32	-32	-28	-26	-24	-22	-19	-20	-17	-15	-15	-22	-18	-22	-26	-26	-28	-47	-37
		Ser			-31	-27	-22	-22	-21	-21	-25	-31	-34	-31	-31	-31	-33	-29	-31	-29	-31	-29	-22	-22	-23
		Thr			-25	-25	-22	-21	-22	-28	-25	-23	-22	-23	-23	-23	-24	-25	-28	-26	-25	-22	-20	-19	-20
		Asp			-21	-21	-26	-30	-30	-35	-30	-30	-30	-26	-30	-21	-21	-26	-26	-26	-26	-21	-20	-20	-18
		Glu			-24	-25	-25	-28	-36	-25	-25	-26	-26	-32	-47	-47	-37	-37	-32	-28	-31	-31	-36	-36	-36
		Asn			-17	-15	-17	-17	-19	-23	-34	-39	-39	-39	-39	-30	-28	-17	-18	-19	-20	-23	-23	-25	-20
		Gln			-20	-23	-29	-29	-34	-45	-29	-35	-35	-35	-35	-26	-29	-23	-29	-29	-29	-26	-30	-29	-29
		Lys			-36	-41	-51	-36	-36	-42	-37	-33	-33	-31	-29	-31	-29	-31	-33	-26	-30	-30	-30	-30	-26
		His			-42	-42	-23	-27	-18	-14	-13	-14	-20	-27	-42	-42	-42	-32	-27	-27	-27	-32	-42	-42	-42
		Arg			-22	-18	-20	-18	-20	-25	-29	-34	-34	-34	-34	-24	-24	-20	-18	-18	-18	-22	-22	-25	-25
		Phe			-26	-26	-42	-42	-43	-43	-43	-27	-27	-24	-24	-21	-34	-32	-32	-43	-43	-32	-32	-23	-20
		Tyr			-35	-29	-30	-26	-26	-30	-30	-35	-45	-30	-24	-24	-24	-24	-30	-19	-21	-19	-19	-18	-21
		Trp			-29	-39	-39	-29	-39	-24	-24	-29	-24	-24	-24	-24	-24	-24	-39	-40	-39	-39	-39	-39	-39
		Cys			-21	-25	-30	-25	-30	-25	-25	-25	-25	-25	-19	-19	-19	-22	-22	-30	-30	-30	-30	-30	-21
		Met			-34	-34	-34	-35	-35	-35	-35	-35	-35	-35	-19	-19	-19	-19	-22	-34	-35	-35	-35	-34	-34
		Pro			-17	-21	-19	-24	-27	-33	-27	-27	-33	-43	-43	-28	-24	-21	-19	-15	-19	-27	-27	-27	-33

Group IV displays a maximum for a value of m less than 0 and a minimum for a value of m greater than 0.

Group V displays no maxima or minima. Although some amino acid residues show rather mixed group characteristics, the effects are usually well defined. An amino acid will usually show entirely different results depending on the conformation under consideration. Assignments of residues are listed in Table 3.

Right-hand α -helical conformation (H_1)

Figs. 3 and 4 represent some of the data in graphic form. The horizontal line represents the estimate of $I(\$^4_{j+m} = \{X\}:\{\bar{X}\})$, which may be regarded as the

corresponding plot for the average type of amino acid residue.

$E[I(\$^4_{j+m} = \{X\}:\{\bar{X}\})|D]$ is -1.2 nats for conformation H_1 and -2.5 nats for P_2 for all values of m when $-10 \leq m \leq +10$. With values measured relative to this line as the abscissa, the plots represent:

$$\begin{aligned} E[I(\$^4_{j+m} = \{X\}:\{\bar{X}\}; R_j)] \\ = E[*(\$^4_{j+m} = \{X\}:\{\bar{X}\}; R_j) | D] \\ - E[I(\$^4_{j+m} = \{X\}:\{\bar{X}\}) | D] \end{aligned} \quad (4)$$

[cf. eqns. (2) and (3)]. The I function on the left-hand side of the equation is employed because it reflects the way in which any residue differs in behaviour from the average type of residue. This set level of significance is chosen by inspection and it is possible

Fig. 3. Directional plots of $I_D = E[I(\$^4_{j+m} = \{X\}:\{\bar{X}\}; R_j) | D]$ versus residue spacing, m , for residues in the H_1 conformation. Ordinate scale: 1 division = 2 decinats. (a) Proline; (b) methionine; (c) tyrosine; (d) glycine. ●, Information that the residues carry about their own conformation. ○, Values of I_D calculated for values of m , $10 \geq m \geq -10$. Curves are drawn to fit the points with a maximum deviation of 3 decinats (see the text).

Fig. 4. Directional plots of $I_D = E[I(\{X\}_{i+m} = \{X\}; \{\bar{X}\}; R_j) | D]$ versus residue spacing, m , for charged residues in the H_1 conformation

Ordinate scale: 1 division = 2 decinats. (a) Glutamic acid; (b), lysine; (c) aspartic acid; (d) histidine; (e) asparagine; (f) arginine. For definition of symbols see Fig. 3.

to fit a smooth curve through the points obtained so that no point deviates by more than 0.3 nat from it. This rule is used in all the plots in the present paper and is in good agreement with the 'noise' level found previously by using a different information statistic (Robson & Pain, 1972).

Group I (maximum at $m = 0$) consists of aliphatic hydrophobic residues not branched at the β -carbon, namely alanine, leucine, methionine and cysteine. Methionine (Fig. 3) is a typical member of the group. The magnitude of its helix-forming effect can be compared with the directional helix-breaking effect

of proline (Fig. 3). The inability of proline to form a helix on its *N*-terminal side is well known, the amide moiety of the peptide group being blocked by covalent interaction with the side chain.

Group II (minimum at $m = 0$) contains glycine and hydroxyl-containing amino acids serine, threonine and tyrosine. Glutamine is a marginal member, on the borderline with group IV. Threonine is skewed to give a minimum at $m = -3$ and hence is a borderline case with group III. Glycine and tyrosine (Fig. 3) are typical members with approximately equal effects.

Group III contains proline (Fig. 3), tryptophan (Fig. 8), aspartic acid and glutamic acid (Fig. 4), and the non-aromatic hydrophobic residues branched at the β -carbon. It is probable that the origin of the effect for the acidic amino acid side chains, like proline, lies in a masking of one of the three amide moieties of the peptide units in the *N*-terminal face of a helical region, so preventing development of the helix in the *N*-terminal direction. It is tempting to postulate a proton-charge-transfer complex between the acidic side chains and the amide moiety. The marked difference between plots for asparagine (Fig. 4) and glutamine and the fact that the positively charged amino acids are such strong members of group IV, are consistent with this hypothesis. Against it, however, examination of the X-ray data seems to indicate that by no means all the aspartate and glutamate side chains are in the right conformation for the formation of such a complex, and it may be that such complexes play a role 'sealing' the helix boundaries mainly during the folding of the protein (Robson & Pain, 1972). The reason that amino acids branched at the β -carbon belong to the same group must lie in the more indirect steric interactions (see Robson & Pain, 1971). It is noteworthy that threonine, which is both a hydroxy amino acid and branched at the β -carbon, is on the borderline between group II (containing hydroxy amino acids) and group III. However, none of these effects is as strong as those involving the acidic amino acids. The effect owing to tryptophan is, however, fairly strong and will be discussed below.

Group IV contains the positively charged amino acid residues lysine, histidine and arginine, the

aromatic amino acid phenylalanine, and asparagine (Fig. 4). Glutamine is a marginal member on a borderline with group II. The origin of the effect associated with the positively charged side chains may be the result of a proton-charge-transfer complex between the basic side chain and one of the three carbonyl moieties on the *C*-terminal face of the α -helix. Otherwise, the arguments for and against such an interaction are entirely analogous to those involving the acidic side chains of group III. The presence of phenylalanine in this group suggests that steric effects can have an important role to play.

Group V, intended to contain those amino acids with no significant effects, is now found to be empty. This discrepancy with previous results (Robson & Pain, 1972), where tyrosine, tryptophan, cysteine and methionine were found to belong to such a group, is almost entirely due to the improved data and to the more stringent definitions of α -helix. An amino acid could be found in this group if there were insufficient data to show a significant effect. Amino acids such as valine, leucine and threonine show effects that are considerably less marked than those found for other members of their groups, but which are still significant by the criterion used here.

β -Pleated-sheet conformation (P_2)

Similar groupings can be made in the case of β -pleated sheet, although they are less easily distinguished (see Table 3). The plots have, in general, a much more complicated profile with some sharp peaks and troughs owing to runs of residues in the P_2 conformation being frequently interrupted by hairpin turns and other local distortions (Robson & Pain, 1974b). These are often associated with the turning back of the chain on itself to form pleated-sheet or ribbon structures.

Group I consists mainly of the hydrophobic amino acids alanine, valine, isoleucine, phenylalanine and tryptophan (Fig. 5). Threonine also belongs to this group. The other hydrophobic amino acids tyrosine, cysteine and methionine all favour the P_2 conformation but, as typified by Fig. 6, the maxima are skewed to $m \approx 2$, placing them on a borderline with group III, together with lysine.

Group II contains glycine, proline and the remain-

Table 3. Assignment of amino acid residues to directional groups on the basis of the profile of their directional plots

For definition of groups see the text.

	Group ... I	II	III	IV
H_1 conformation	Ala, Leu, Met, Cys	Gly, Ser, (Thr), Tyr	Val, Ile, Trp, (Thr), Asp, Glu, Pro	Lys, His, Arg, Gln, Asn, Phe
P_2 conformation	Ala, Val, Ile, Phe, Trp, Thr, (Tyr), (Lys), (Met), (Cys)	Gly, Glu, Ser, Arg, Asn, His, Pro	Asp, Gln, (Tyr), (Cys), (Met), (Lys)	Leu

Fig. 5. Directional plots of $I_D = E[I(\{X\}_{+m} = \{X\} : \{\bar{X}\}; R)]D$ versus residue spacing, m , for residues in the P_2 conformation. Ordinate scale: 1 division = 2 decinats. (a) Serine; (b) alanine; (c) arginine; (d) threonine. For definition of symbols see Fig. 3.

ing polar amino acids with the exception of aspartic acid and glutamine. Serine and arginine (Fig. 5) have symmetrical minima about $m=0$. Glycine, histidine and glutamic acid, as typified by glycine (Fig. 7), possess very sharp and very deep minima which are, however, skewed so that these amino acids tend to occur at the N -terminal end of discontinuities in β -pleated-sheet structures. The same applies to proline and to a greater extent, asparagine (Fig. 7), except that these amino acids tend to occur at C -terminal ends of discontinuities in runs of P_2 conformations. The members of this group, particularly glycine, asparagine, proline, glutamic acid and histidine with their sharp deep maxima, are associated with various kinds of turn structure (Robson & Pain, 1974b) and with reverse turns in particular (Crawford *et al.*, 1973). The detailed mechanism of this association is not clear except in the case of glycine, which is sterically favoured in the G and L conformations so common in turn

structures. However, a minimal requirement for hydrogen-bonding side chains has also been noted in turns that contain G or L conformations (Robson & Pain, 1974b).

Group III contains aspartic acid and glutamine, which both tend to occur at the C -terminal end of discontinuities in runs of P_2 conformation. Again, these amino acids are abundant in turn structures in general (Robson & Pain, 1974b) and in reverse turns in particular (Crawford *et al.*, 1973). Tyrosine (Fig. 6), cysteine and methionine are marginal members of this group, tending to occur within and towards the N -terminal of a run of pleated-sheet conformations.

Leucine should correctly be placed in group IV and this amino acid is the only one that clearly prefers to occur within but towards the C -terminal of a run of pleated-sheet conformation.

As with the α -helical conformation, there are no members of group V.

Fig. 6. Directional plots of $I_D = E[I(\mathcal{S}_{j+m}^4 = \{X\}; \{\bar{X}\}; R_j) | D]$ versus residue spacing m , for isoleucine (a) and tyrosine (b) in the P_2 conformation

Ordinate scale: 1 division = 2 decinats. For definition of symbols see Fig. 3.

Fig. 7. Directional plots of $I_D = E[I(\mathcal{S}_{j+m}^4 = \{X\}; \{\bar{X}\}; R_j) | D]$ versus residue spacing m , for glycine (a) and asparagine (b) in the P_2 conformation

Ordinate scale: 1 division = 2 decinats. For definition of symbols see Fig. 3.

Discussion

These results quantify the various tendencies for different conformations to occur in runs and confirm that there are powerful correlations between amino acid sequence and conformation. Moreover, it is clear that individual amino acids not only have a role in determining the preferred conformation but also modify the preferred lengths of runs of that conformation. The latter phenomenon has been demonstrated by plotting values of $E[\star(\$_j = \{X\}:\{\bar{X}\}; R_j)|D]$, the information given by the residue R_j as to its occurrence in a string of n residues having the conformation X , against n (Fig. 2) and by measuring the intercepts at $n = 1$ and the slopes for each of the 20 amino acids for four conformations (Table 1). All other conformations on the domain diagram (Fig. 1) have been analysed, but residues in these conformations show none of the statistical co-operativity discussed below. The slopes vary with the domain because different conformations tend to have different string lengths, e.g. those in the P_2 pleated-sheet conformation are shorter on average than those in conformation H_1 , the α_1 -helix. Within these lengths for each conformation, however, each residue R_j has its own effect on the slope.

The slope for a residue R_j indicates the statistical co-operativity and provides a measure of the information required to bring an undefined residue up to a string of residue conformations X given that R_j is located within the residue string and in the conformation X . Conversely, the slope also measures the required X -forming power of the residue to be brought from conformation $\bar{X}$ to join the X -string containing R_j . Taking X as the helical H_1 conformation, for example, the information required is high if the probability of extending the helix is a relatively rare event and a residue of high helix-forming power would be required to extend the helix. This will be the case when the occurrence of R_j in a helix has a fragmenting effect on the helix, i.e. occurs near a helix boundary. Because of this, the slope can further be seen as a measure of the fragmenting power of R_j .

These definitions hold only when R_j is itself in the helix, which may or may not be a frequent event. The slope has nothing to say concerning the preference of R_j for the helical (or, more generally, X) conformation. This latter is given by the value of the ordinate at $n = 1$, which is a measure of the single residue or intra-residue information for R_j . To take an example from Table 1, glycine does not like to be helical, but when forced to be so it favours helical neighbours. Proline, on the relatively few occasions that it occurs in runs of conformation H_1 , exerts a strong fragmenting effect, implying that extension of the helix in which it occurs would require addition of a residue having very high helix-forming power.

Analogy with helix-coil transition-theory parameters

Evidence has been presented which supports the argument that the formation of strings of residues in regular conformations, such as the α -helix, is important in the formation of starting conformations for the folding of globular proteins (Robson & Pain, 1973, 1974a). Therefore it would be useful to be able to correlate the measured parameters in Table 1 with parameters from theories of helix-coil transitions. Several theories exist for studying helix-coil transitions including notably that of Lifson & Roig (1961). With the use of the notation of Zimm & Bragg (1959), two main parameters s and σ are proposed, the statistical weights of the coil residues, of the helical residues with a full complement of helix hydrogen bonds, and of boundary residues without a full complement of helix hydrogen bonds being 1, s and $\sigma^{1/2}$ respectively (cf. Poland & Scheraga, 1970). A link may be seen between σ , which varies with the molecular co-operativity of the helix-coil transition, and the ease or otherwise of bringing a residue c' in the run $H H H c' c c$ into the H conformation. The values in Table 1 cannot be used to order residues in terms of the contribution they can make to the parameter σ since the slope parameter of Fig. 2 applies to the residue R_j located at some point within the helix. The experimental molecular co-operativity for extending this helix would be the value $dE[\star(\$_j = \{X\}:\{\bar{X}\}; R_j)|D]/dn$ contributed by all the residues in the vicinity of the helix-coil boundary.

The parameter s is more clearly linked to the value of the intercept, and can be shown to be equal to $\exp[\star(\$_j = \{X\}:\{\bar{X}\}; R_j)]$ when $n = 1$ and $\sigma = 1$. The value at the intercept strictly contains information derived from all residues in an H_1 conformation including those which are not in an α_1 -helix (Robson & Pain, 1974a). Therefore the ordinate value corresponding most closely to s will be that at $n = 4$, i.e. the values given by the filled circles on the directional plots at $m = 0$, when $\sigma \neq 1$.

Thus parameters s are related to the information parameters and are expected to increase monotonically with them.

The results of the analysis presented in the present paper may be used to study and predict the transitions of co-operative structures in globular proteins. Lewis & Scheraga (1971) have taken the preference measures of Pain & Robson (1970) [similar to those expressed here by the ordinates for conformation H_1 (Table 1)] and used them to assign s values to all residues, by using the few experimentally determined values of s . Making the assumption that a coil to helix transition is a primary event in the folding of a globular protein, they then calculated helix-probability profiles in which the regions of maximum helix-probability frequently correspond to the helical

regions observed by X-ray crystallography. The good agreement shown (Robson & Pain, 1973) between the approaches of Lewis & Scheraga (1971) and the information approach (which does not depend for its rationale on this or any other specific model of folding) may be regarded as strong indirect evidence in favour of the initial assumption.

A more direct application of the Zimm & Bragg (1959) theory of helix-coil transitions (see Poland & Scheraga, 1970, for a review) would involve the estimation in a similar way of s parameters from their monotonic relation to the intercept at $n = 4$.

The need for such a combination of experimental and theoretical approaches arises from the difficulties in obtaining s and σ experimentally. First, a homopolymer can only be used if it is soluble in water, forms the required conformation in water and undergoes a transition between the required conformation and the time-random coil conformation in the accessible range of temperatures. Secondly, it is difficult to obtain the required parameters for the formation of a conformation which is not the most preferred, such as for β -pleated sheet as opposed to α -helix. In fact, the probability-profile approach, based on at least some experimental parameters, has essentially been applied only to the α -helical conformation. Thirdly, although these difficulties can in principle be circumvented by using the amino acid of interest as a 'guest residue' introduced at random into a host sequence with appropriate properties (Scheraga *et al.*, 1971), one must be sure that the perturbations of the s and σ of the host do not arise from neighbour interactions that are peculiar to that combination of host and guest.

There are special difficulties in attempts to apply Zimm-Bragg type of theory to the calculation of probability profiles for β -pleated sheet and many other non-helical conformations. The theory applies formally only to local interactions, whereas pleated-sheet structures involve hydrogen bonds between other chains or distant regions of the same chain. We have already discussed (Robson & Pain, 1974a) the importance of local effects in determining β -pleated-sheet conformation. In particular, hydrogen bonds are generally formed to residues in the pleated-sheet conformation whether or not a pleated-sheet or ribbon structure is actually present. However, this applies to the situation in the stably folded protein. In the information-theory approach to prediction of β -pleated-sheet conformations the information measures are derived from native proteins and hence the same kind of system as that to which they are to be applied. The extension of Zimm-Bragg theory to such predictions, on the other hand, may be invalidated because the analogous measures are derived from an artificial experimental system in which long-range interactions might have

rather different consequences. The applicability of the Zimm-Bragg theory to the prediction of β -pleated sheet needs therefore to be rigorously proved.

It would be possible to carry the approach of Lewis & Scheraga (1971) to its limit and use information theory not to supplement experimental assignments but to provide all assignments. At this limit, however, the predictions would not represent an application of the Zimm-Bragg theory, but would merely make use of the mathematical procedures of Zimm & Bragg (1959) as a probability-theory approach analogous to the information-theory approach. More specifically, any predictive success enjoyed by application of the Zimm-Bragg mathematics to β -pleated-sheet predictions in these circumstances would not provide evidence for the formation of β -pleated-sheet conformation as a primary event in the folding of globular proteins from a time-random coil.

Relation to previous studies of directional effects

The variation of the quantity $\star(\$^4_{+m} = \{X\}; \overline{\{X\}}; R_j)$ with m from $m = -10$ to $+10$ provides the data for directional plots (Table 2 and Figs. 3-8), which quantify the effect a residue has on the conformation of a neighbour at a specific distance and in a specific (N - or C -terminal) direction. The results confirm and extend the general findings of Robson & Pain (1972), who used a slightly different information statistic and studied the α -helix only. Any differences are presumably due to improved data, since the previous work made use of helical regions published without detailed ϕ - ψ angles. In the present work, much clearer directional plots are produced without any kind of artificial smoothing, and residues then described as having flat plots, within the level of significance used, now show significant features.

The present results fully confirm the previous findings concerning the marked tendency for negatively charged proton acceptors (glutamic acid and aspartic acid) to occur at the N -terminal end, and for positively charged proton donors (lysine, histidine, arginine) to occur at the C -terminal end of helical regions. On the other hand, the present analysis gives, for other amino acid residues, much clearer plots that do not always reproduce the earlier indications (Robson & Pain, 1972) of asymmetry about $m = 0$. Presumably the discrepancy arises from changes in the assignment of helical boundaries that have occurred in going from the old definitions of helix to the more accurate and detailed boundaries given in terms of ϕ - ψ angles. Glycine emerges as a helix-breaker with more or less equal effect in each direction. However, the co-operativity results show that when forced into a helical conformation it will propagate the helix,

and therefore glycine is better termed a non-initiator of helix.

The residues with a hydroxyl side chain (serine and threonine) are also more or less symmetrical about $m = 0$ and are helix-breaking. By analogy with the positively charged donor and negatively charged acceptor side chains, it might be concluded that an ambivalent moiety such as a hydroxyl group would enjoy both an *N*- and *C*-terminal helix boundary, resulting in a symmetrical helix-break or no helix at all. The fact that threonine is skewed so as to tend slightly to the *N*-terminal end of helices does not contest this hypothesis, since this skew is a property of other side chains with branched β -carbon atoms (valine and isoleucine). On the other hand, the result for asparagine does suggest that there might be other important factors involved since this residue has a marked dislike for helix on its *C*-terminal side. The total effect of a helix-breaking amino acid will depend not only on its helix-destabilizing effect, but on the stabilizing influence it has on other conformations. In other words, it is only the relative energies of the conformations that are meaningful. In this case, asparagine has a very marked preference for turn structures (Robson & Pain, 1974b) as revealed in the present paper by a discontinuity in a run of residues in the P_2 pleated-sheet conformation.

Although not always so clear cut, the results for pleated-sheet conformations show that certain residues have marked directional effects in this structure as in the α -helix. The hydrophobic amino acids in general have a conformation P_2 -forming effect on their neighbours. The initiation of discontinuities in β -pleated-sheet structures is a property in particular of glycine, asparagine, proline, glutamic acid and histidine. These results complement the analysis of the role of residue pairs in the various types of turn and kink which can interrupt runs of P_2 conformations.

Relevance to distortions of co-operative structures

Some preliminary observations on the distortion of α -helix and β -pleated sheet are relevant to the interpretation of the results presented here. Distortions of α -helix or β -pleated sheet are frequently found which have not, in general, been differentiated from classical helix and pleated-sheet structures. Runs of residues in the P_2 conformation are frequently interrupted by residues in the P_1 or P_3 conformation, and runs of residues in the H_1 conformation are similarly interrupted by residues in the H_2 conformation. The ability to treat distortions separately (Robson & Pain, 1974a) is valuable because it demonstrates a correlation between the distortions and the amino acids occurring in the region of the distortions. For example, proline is very frequently associated

with the P_3 conformation and this is the principal reason for its classification as a breaker of a run of P_2 conformations. Such a distortion is usually associated with a turn structure (Crawford *et al.*, 1973; Robson & Pain, 1974b).

Of particular interest are the correlations between amino acid sequence and distortions of the right-hand α -helix. The conformation H_2 , which represents this distortion, has variously been defined as an element of a 3_{10} -helix, of α_{11} -helix (Nemethy *et al.*, 1967), or of type I or type III reverse turns (Crawford *et al.*, 1973), although these are clearly all closely related conformations. The particular significance of the α_{11} -helix is that it can explain the tendency for α -helices to end in the H_2 conformation (Robson & Pain, 1971). Distortion of the peptide bond associated with the α_{11} conformation allows bifurcate helical hydrogen bonding in the *N*-terminal direction, but disallows stable helical hydrogen bonding in the *C*-terminal direction (Nemethy *et al.*, 1967). Because of this tendency, a residue which tends to occur towards the middle of a helical region, such as alanine, would be expected to have a symmetrical helix-forming H_1 plot, but an asymmetric H_2 plot having the form of:

$$\begin{aligned} E[I(\$)_{+m} = \{H_2\} : \{\overline{H_2}\}; R_j] / D] \\ = -dE[I(\$)_{+m} = \{H_1\} : \{\overline{H_1}\}; R_j] / dm \end{aligned}$$

as shown by Fig. 8.

Despite this tendency for conformation H_2 to occur at the *C*-terminal end of α -helices, there are many instances of H_2 conformations occurring within helices. In an attempt to explain the features for the H_1 plot of tryptophan, corresponding plots were generated for the other conformations. The dislike of tryptophan for conformation H_1 on its *N*-terminal side was found to result from a liking for the H_2 conformation on this side (Fig. 8). A plot constructed for tryptophan for the conformation H , which is the union of H_1 and H_2 , would be helix-favouring and symmetrical around $m = 0$. This effect is reflected by the tendency of tryptophan to be helical, but to occur at the end of a run of H_2 conformations. It is therefore detected at the *C*-terminal side of type I or type III reverse folds and at the *C*-terminal side of undifferentiated reverse turns (Crawford *et al.*, 1973).

Analysis of side-chain conformation in proteins of known conformation is made less useful by the fact that interactions responsible for stabilizing, say, α -helix and helix boundaries (Robson & Pain, 1971), might not persist once the helix is fixed by tertiary interactions. Analysis of steric interactions which may arise during the folding process are occasionally misleading. For example, residues branched at the β -carbon should prefer the α_{11} conformation, at least for some side-chain conformations. Threonine in particular has a marked tendency to occur in the α_{11} conformation (Robson & Pain, 1971). However,

Fig. 8. Directional plots of $I_D = E[I(S_{j+m}^A = \{X\}; \{\bar{X}\}; R_j) | D]$ versus residue spacing, m : comparison of H_1 and H_2 conformations for alanine and tryptophan

Ordinate scale: 1 division = 2 decinats. H_1 conformation: (a) alanine; (b) tryptophan. H_2 conformation: (c) alanine; (d) tryptophan. For definition of symbols see Fig. 3.

despite the fact the α_{II} conformations occur most frequently at the C-terminal ends of helices, these amino acids do not belong to group IV but to group III with the opposite symmetry. In fact, there are several strong effects involving the α_{II} conformation, as exemplified by tryptophan, in which the H_1 conformation is not disfavoured on the C-terminal side. The implication is that there are frequently two simultaneous mechanisms for helix disruption. One of these involves forcing an α_{II} conformation

(see Pain & Robson, 1972), which should result energetically in helix disruption because of the distorted hydrogen-bonding system. The other must be an essentially stabilizing influence which is strong enough to over-ride the expected consequences of a α_{II} conformation. The alternative explanation would be that the α_{II} conformation is a consequence, rather than a cause, of the C-terminal boundary of a α -helix, and that it has little to do with the coding of the α -helix.

Note Added in Proof (Received 10 June 1974)

Chou & Fasman (1974) point out some differences between their frequencies of helical boundary and central residues and the directional effects published earlier (Robson & Pain, 1972). The assignments in the present paper, based on qualitatively and quantitatively better data, show the frequencies given by Chou & Fasman (1974) to be compatible with our detailed directional effects.

This work was supported by a grant from the Medical Research Council. The expert assistance of Miss Moira Bain is gratefully acknowledged.

References

- Anfinsen, C. B. (1967) *Harvey Lect.* **61**, 95–116
Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 211–222
Crawford, J. L., Schellman, C. & Lipscomb, W. N. (1973) *Proc. Nat. Acad. Sci. U.S.* **70**, 538–542
Guzzo, A. V. (1965) *Biophys. J.* (1971) **5**, 809–822
Havsteen, B. H. (1966) *J. Theor. Biol.* **10**, 1–10
Kotelchuk, D. & Scheraga, H. A. (1968) *Proc. Nat. Acad. Sci. U.S.* **61**, 1163–1170
Kotelchuk, D. & Scheraga, H. A. (1969) *Proc. Nat. Acad. Sci. U.S.* **62**, 14–21
Lewis, P. N. & Scheraga, H. A. (1971) *Arch. Biochem. Biophys.* **144**, 576–583
Lifson, S. & Roig, A. (1961) *J. Chem. Phys.* **34**, 1963–1974
Nemethy, G., Phillips, D. C., Leach, S. J. & Scheraga, H. A. (1967) *Nature (London)* **214**, 363–365
Pain, R. H. & Robson, B. (1970) *Nature (London)* **227**, 62–63
Pain, R. H. & Robson, B. (1972) *Proc. Eur. Congr. Biophysics 1st* **1**, 33–37
Phillips, D. C. (1967) *Proc. Nat. Acad. Sci. U.S.* **57**, 484–495
Poland, D. & Scheraga, H. A. (1970) *Theory of Helix-Coil Transitions in Biopolymers*, Academic Press, New York and London
Prothero, J. W. (1966) *Biophys. J.* **6**, 367–370
Ptitsyn, O. B. (1969) *J. Mol. Biol.* **42**, 501–510
Quirocho, F. A. & Lipscomb, W. N. (1971) *Advan. Protein Chem.* **25**, 1–78
Robson, B. (1974) *Biochem. J.* **141**, 853–867
Robson, B. & Pain, R. H. (1971) *J. Mol. Biol.* **58**, 237–259
Robson, B. & Pain, R. H. (1972) *Nature (London) New Biol.* **238**, 107–108
Robson, B. & Pain, R. H. (1973) in *Conformation of Biological Molecules and Polymers, 5th Jerusalem Symp.* (Bergmann, E. D. & Pullman, A., ed.), pp. 161–172, Academic Press, London and New York
Robson, B. & Pain, R. H. (1974a) *Biochem. J.* **141**, 869–882
Robson, B. & Pain, R. H. (1974b) *Biochem. J.* **141**, 899–904
Scheraga, H. A., von Dreele, P. H. & Poland, D. (1971) *Macromolecules* **4**, 396–407
Tanford, C. (1968) *Advan. Protein Chem.* **23**, 121–282
Zimm, B. H. & Bragg, J. K. (1959) *J. Chem. Phys.* **31**, 526–535