Proc. Nat. Acad. Sci. USA
Vol. 70, No. 5, pp. 1473–1477, May 1973

The Influence of Nearest-Neighbor Amino Acids on the Conformation of the Middle Amino Acid in Proteins: Comparison of Predicted and Experimental Determination of β-Sheets in Concanavalin A

ELVIN A. KABAT AND T. T. WU

Departments of Microbiology, Human Genetics and Development and Neurology, Columbia University; The Neurological Institute, Presbyterian Hospital, New York, N.Y. 10032; and the Departments of Engineering Sciences and Physics, Northwestern University, Evanston, Illinois 60201

Contributed by Elvin A. Kabat, March 12, 1973

ABSTRACT **A 20 × 20 table of tripeptides has been compiled that may be used to locate β-sheet breaking and α-helix breaking residues in proteins. It is based on the definition of an α-helical and a β-sheet domain on the (ϕ, ψ) map based on the occurrences of α-helices and β-sheets in 12 known proteins whose sequence and three-dimensional structure have been determined. Each entry in the 20 × 20 table lists three numbers, the frequency of occurrences of the middle amino acid (n) in relation to its nearest neighbors ($n - 1$) and ($n + 1$) in the α-helical domain, the β-sheet domain and outside these regions. The regions between two β-sheet-breaking residues would be permissively β-sheet regions. The sequence of concanavalin A has been examined in this manner and of the 13 β-strands defined by x-ray crystallography, 10 were in agreement with the permissively β-sheet regions and, in the remaining three, β-sheet-breaking residues were the third in one, and the third, fourth, and fifth residues in another, and the sixth residue in the third from the beginning of the β-strands. The findings provide strong support for the role of nearest-neighboring amino acids in determining secondary structure of proteins.**

The influence of nearest-neighbor amino acids, ($n - 1$) and ($n + 1$), on the conformation of any amino acid (n) in a protein molecule has been extensively studied (1–4). Two approaches have been developed, both empirical and based on data on the (ϕ, ψ) angles of the middle amino acid (n) in tripeptides ($n - 1$) (n) ($n + 1$) compiled from proteins whose sequence and whose three-dimensional structures were known. The first (1, 4) locates an α-helical and a β-sheet domain on a (ϕ, ψ) map based on the (ϕ, ψ) angles of residues in these proteins known to be in an α-helix or in a β-sheet, excluding the end residues. The sequences of these known proteins are then examined as tripeptides e.g., 1, 2, 3; 2, 3, 4; 3, 4, 5, etc., with respect to the nearest-neighboring amino acids ($n - 1$) and ($n + 1$) to evaluate their influence on the (ϕ, ψ) angles of amino acid (n). These effects can be summarized in a 20 × 20 table of ($n - 1$) and ($n + 1$) that lists the number of instances in which the middle amino acid of a tripeptide occurred in a right-handed α-helical conformation, in a β-sheet conformation, and in neither, as observed in proteins with known tertiary structures (4). For a given pair of nearest-neighbor amino acids these three numbers may provide indications of their tendency to disrupt α-helices and β-sheets (4). Thus for the nearest-neighbor tripeptides, Tyr-( )-Tyr and Lys-( )-Leu, the incidences of values in the α-helical and β-sheet domains and in neither were 0, 2, 3 and 11, 0, 6 and the residue in the middle of each tripeptide would be designated as an α-helix-breaking or a β-sheet-breaking residue, respectively. The sequences between two α-helix-breaking or two β-sheet-breaking residues are termed permissively helical or permissively β-sheet regions. The table, which considers only the influence of amino acids ($n - 1$) and ($n + 1$) without regard to the nature of amino acid (n), may be used with but a single sequence.

This 20 × 20 table has been used successfully with several proteins, e.g., variable regions of light and heavy chains of immunoglobulins whose three-dimensional structures have not yet been determined and with cytochrome c, cytochrome b_5, thermolysin, etc., to predict the absence of α-helical segments at specific regions (1, 4). However, due to limited data on proteins with extensive β-sheet segments, the predictive power of this table on the absence of β-sheets has not been fully studied, earlier observations having been made only on papain (4).

The second approach (2, 3) uses the exact (ϕ, ψ) angles of (n) in tripeptides in known proteins or tries to approximate them if exact values are not available, and attempts to predict a set of (ϕ, ψ) angles for other proteins whose sequence is known. These (ϕ, ψ) data are expressed in twenty 20 × 20 tables to encompass the 8000 possible tripeptides. For the selection of a set of (ϕ, ψ) angles by this method, a series of homologous proteins of known sequence is required. Human κ immunoglobulin light chains (2) and 18 cytochromes c from 18 species have been examined (3).

Recently, the primary (5) and tertiary (5, 6) structures of concanavalin A, a metalloprotein lectin with receptor sites most specific for α-D-mannopyranosyl residues (7) and which precipitates with various polysaccharides and glycoproteins and (8–11) agglutinates (8, 12, 14) and is mitogenic (15) for certain cells, have been elucidated. It is made up of many β-sheet regions, as deduced from the x-ray data (5, 6). The present study locates the β-sheet-breaking residues in concanavalin A based on its primary amino-acid sequence using the first procedure and relates the position of these residues to the observed β-sheets.

EXPERIMENTAL DATA

The 20 × 20 table (4) was constructed from the known tertiary structures and sequences of 11 proteins: Sperm whale myoglobin (16), the α- and β-chains of horse oxyhemoglobin

TABLE 1. *Frequency of occurrences of various conformations of amino acids at position (n) based on the*

(n+1) \ (n−1)	TRP	ILE	TYR	PHE	PRO	LEU	VAL	LYS	MET	CYS
TRP	1•0•0	2•0•1	1•0•1	0•0•0	0•0•1	0•0•0	3•0•1	0•0•1	0•0•0	0•0•0
ILE	2•1•0	2•1•1	0•0•1	0•0•1	1•1•0	2•3•4	1•1•1	4•2•4	0•2•0	0•0•1
TYR	1•0•0	2•0•1	0•2•3	0•1•4	2•1•0	0•2•3	1•1•3	2•2•1	0•0•1	0•1•0
PHE	0•0•0	0•0•0	1•1•1	1•1•1	2•0•3	1•1•1	3•1•2	3•3•3	0•1•1	0•0•0
PRO	1•0•0	1•0•1	0•0•7	0•1•2	0•0•0	1•1•2	0•0•7	1•2•4	0•2•0	0•2•0
LEU	1•1•1	2•3•0	0•2•4	2•2•2	2•2•3	8•3•2	1•1•3	11•0•6	2•0•1	3•1•2
VAL	0•1•1	2•5•2	2•0•3	1•0•1	3•4•2	7•1•7	2•6•2	6•1•3	0•0•1	0•0•2
LYS	2•0•2	2•0•3	3•0•1	5•0•4	3•0•2	12•2•3	3•5•3	6•1•6	2•1•0	2•0•2
MET	0•0•0	0•1•0	0•0•0	0•0•0	0•0•0	1•0•0	0•0•0	1•1•0	0•0•0	0•0•0
CYS	1•0•1	0•0•1	0•0•0	0•0•0	0•0•1	2•1•1	1•1•1	0•0•0	1•1•0	0•0•0
ALA	3•0•1	3•4•2	1•0•2	2•0•4	6•2•1	9•1•4	19•7•7	8•6•1	2•0•0	2•0•0
ARG	0•0•1	0•0•1	0•0•2	2•1•1	1•0•0	7•0•2	2•0•0	3•0•3	1•0•0	0•0•0
THR	1•0•1	1•1•2	0•5•3	2•1•3	3•2•2	3•0•1	4•4•2	3•3•2	0•0•0	0•1•4
SER	0•0•1	2•2•9	0•2•2	6•2•0	1•0•6	6•2•7	6•1•5	1•1•5	2•0•2	0•0•4
GLY	1•0•0	1•1•3	0•1•9	0•1•3	2•1•2	4•2•7	5•2•9	4•2•10	0•0•0	0•3•3
HIS	0•0•0	1•1•0	0•0•0	2•1•1	0•0•1	2•0•1	5•0•1	7•0•3	1•0•0	0•0•0
ASP	0•1•0	6•3•0	1•1•1	2•0•3	0•0•2	3•3•4	2•2•1	0•1•0	0•0•3	0•1•0
ASN	1•1•1	2•1•1	1•1•3	0•1•2	3•1•2	7•0•9	3•1•2	1•2•4	2•0•0	1•0•0
GLU	0•0•2	4•1•1	2•0•1	1•1•2	0•0•1	5•0•3	1•0•2	7•2•1	1•0•0	0•1•3
GLN	2•0•0	1•1•1	0•0•3	1•1•1	0•0•1	4•1•1	4•1•3	0•3•1	0•0•0	1•1•2
TOTAL	17•5•12	34•25•30	12•15•47	27•14•35	29•14•30	84•23•62	66•34•55	68•32•58	14•7•9	9•11•23

The first value in each entry gives the number of occurrences in the α-helical domain, the second those in the β-sheet region, and the third those in neither region [including those on the borders of the α-helical and β-sheet domains as defined (see refs. 1 and 4)].

(Perutz, private communication), lysozyme (17, 18), tosyl-α-chymotrypsin (ref. 19 and private communication), carboxypeptidase A (ref. 20 and private communication), ribonuclease S (ref. 21 and private communication), subtilisin (refs. 22 and 23 and private communication), lamprey hemoglobin (ref. 24 and private communication), staphylococcal nuclease (ref. 25 and private communication), insulin (refs. 26 and 27 and private communication), and horse cytochrome *c* (refs. 28 and 29 and private communication). Since the (ϕ, ψ) values of cytochrome b_5 became available recently (30, 31), they are also incorporated. The (ϕ, ψ) angles for tosyl-α-chymotrypsin have been revised (32) and these values have been substituted (Table 1). The delineation of the α-helical and β-sheet domains have been described (1, 4).

In the tabulation of the α-helical and β-sheet frequencies for the 20 × 20 table, the end residues of the known α-helices and β-sheets were excluded. It is thus not necessarily inconsistent to have a helix-breaking or β-sheet-breaking residue as a terminal residue in an α-helix or β-sheet. It is also evident that there is no *a priori* reason for the permissively helical or permissively β-sheet regions actually to contain α-helices or β-sheets. The test of the method rests on the *absence* of α-helix-breaking and β-sheet-breaking residues inside of known helices and known β-sheets and its usefulness rests on its ability to locate those portions of a sequence in which α-helices and β-sheets might occur.

The primary (5) and three-dimensional (5, 6) structures of concanavalin A have recently been reported. Asx and Glx will be considered both as Asn and Asp and Gln and Glu, respectively. When the nearest-neighboring amino acids are both uncertain, their influence will be omitted.

RESULTS

Table 2 lists the β-sheet-breaking residues as well as α-helix-breaking residues and the observed β-sheets (5, 6) in concanavalin A. A β-sheet-breaking residue is defined as one with no occurrences in the regular β-sheet region (4) and with at least three other occurrences as compiled from the data on the 12 proteins (Table 1). An α-helix-breaking residue, on the other hand, is defined as one with less than 20% occurrences in the α-helical region or with at least three occurrences in the 12 proteins and outside the α-helical domain (Table 1).

Among the 12 β-sheets reported by Edelman *et al.* (ref, 5 and private communication), seven (4–9, 25–30, 73–78. 92–97, 106–116, 140–144, and 209–215) are consistent with the locations of the permissively β-sheet regions. Their recently added short β-strand 36–38 is also consistent. Of the five inconsistent ones, in two (59–66 and 173–177) the hydro-

amino acids at positions (n − 1) and (n + 1) in 12 reference proteins

	(n−1) ALA	ARG	THR	SER	GLY	HIS	ASP	ASN	GLU	GLN	TOTAL
	2,1,0	0,0,4	0,0,1	1,0,3	4,0,1	1,0,0	0,0,0	2,0,0	2,0,0	1,0,0	19, 1, 14
	3,3,2	2,0,1	4,1,3	3,2,4	2,1,1	1,1,0	2,1,0	1,0,2	5,0,1	2,0,2	36,20, 30
	4,1,1	0,0,2	0,2,2	2,0,8	4,3,0	1,1,0	0,0,2	0,1,2	1,0,5	0,0,2	20,18, 40
	3,1,4	3,0,3	0,0,3	1,0,1	1,3,3	1,0,0	2,2,2	0,1,3	5,0,2	1,0,0	28,15, 31
	1,0,1	0,1,1	0,1,2	0,3,7	0,4,5	0,0,1	1,1,0	0,2,1	1,0,2	1,1,1	8,21, 44
	11,3,2	3,0,4	5,2,2	7,2,4	5,2,4	3,0,4	5,1,3	2,0,2	10,0,4	2,3,2	85,28, 55
	12,2,6	0,1,0	4,1,3	6,1,8	5,4,1	3,0,2	5,2,1	3,1,0	4,0,3	2,1,3	67,31, 51
	13,3,5	1,1,2	1,2,8	2,1,4	4,2,4	3,0,4	1,1,1	6,3,2	4,1,1	2,0,2	77,23, 59
	4,1,0	0,0,0	2,0,2	0,0,3	1,1,0	0,0,0	1,1,3	0,0,2	1,0,1	1,0,2	12, 5, 13
(n + 1)	0,0,4	0,0,1	1,2,1	2,2,3	1,0,0	0,1,2	0,0,1	2,0,3	2,0,0	0,0,3	13, 8, 23
	15,0,5	2,0,1	7,3,9	6,2,4	4,5,7	3,0,0	4,1,0	3,0,4	3,1,4	3,1,1	105,33, 57
	2,0,1	1,0,0	2,2,1	0,0,1	0,0,5	3,0,0	2,1,2	1,0,3	0,0,3	1,0,0	28, 4, 26
	3,1,3	2,0,5	2,1,5	6,4,3	2,3,10	2,0,2	2,1,7	1,0,4	2,0,1	3,0,4	42,27, 64
	4,3,9	2,0,1	3,2,8	1,1,13	2,4,15	1,1,1	4,0,8	0,0,4	3,0,1	2,3,2	46,24,103
	8,3,9	0,0,3	2,1,4	5,2,12	1,1,6	3,0,2	1,4,7	1,1,7	4,0,4	0,0,2	42,25,102
	4,0,3	2,0,0	0,1,3	3,1,2	4,2,4	1,0,0	0,0,0	1,3,3	2,0,3	1,0,1	36, 9, 26
	3,1,6	2,2,0	2,0,4	4,2,4	5,2,2	3,3,3	0,0,2	1,0,3	1,1,2	0,1,0	35,24, 40
	1,2,5	0,0,0	1,1,3	4,1,4	1,0,4	0,0,3	1,0,2	4,1,2	1,1,2	0,0,4	34,14, 53
	6,1,6	1,0,0	3,1,4	3,0,2	7,1,4	2,0,4	4,1,1	2,1,0	5,1,1	1,0,2	55,11, 40
	1,1,1	1,0,1	1,1,2	2,2,3	0,0,1	2,0,1	3,0,0	2,0,3	3,0,1	0,0,0	28,12, 26
	100,27,71	22,5,29	40,24,70	57,26,93	53,38,77	32,7,30	38,17,42	32,14,50	39,5,42	23,10,33	816,353,897

gen bonding shows irregular β-structure. Thus, only three out of the 12, i.e., 48–55, 126–132, and 190–199, appear inconsistent with the locations of β-sheet-breaking residues; the discrepancies involve a β-sheet-breaking residue 53 (2, 0, 1) as the fifth residue in the experimentally determined β-strand 48–55; the third residue, e.g., 127 (4, 0, 3) in 125–132 and the fourth and fifth residues, respectively, 192 (1, 0, 2), 193 (2, 0, 4), and 194 (2, 0, 1) in 190–199. The frequency data for residues 53, 192, and 194 are quite sparse and might not be β-sheet-breaking as more data are incorporated.

The values of Hardman and Ainsworth (ref. 6 and private communication) for residues in β-sheets as modified to fit the sequence of Edelman *et al.* (5) (4–10, 35–38, 59–66, 74–79, 90–98, 106–116, 140–146, 171–177, and 209–216) are generally consistent with the permissively β-sheet regions and with the data of Edelman *et al.* (5) as considered above. However, their β-strand 24–30 is not in agreement with β-sheet-breaking residue 25 (3, 0, 2). Their β-strand 90–98 would include the β-sheet-breaking residue 92 (0, 0, 3) as the third residue in the sheet rather than the first. Their β-strands 45–56, 126–132, and 191–199 show the same β-sheet-breaking residues inside the strands as discussed above. Their additional β-strand 154–156 fits well with β-sheet-breaking residues 153 and 157.

The numerous α-helix-breaking residues together with the helical wheel method (33) suggest that there are very few helices. The only possible one, 211–218, that would be supported by the helical wheel method is at the location of one of the observed β-sheets. The only single turn of observed α-helix is at residues 81–85 (5, 6).

Probabilities for β-bends (34, 35) have also been calculated by use of the table based on six proteins (35). The highest ones are located at 19–22 (8.38×10^{-4}), 165–168 (7.06×10^{-4}), and 222–225 (4.99×10^{-4}). From the actual model, however, they are seen to be a bend involving five amino acids and one of the metal ions, an extended chain, and a distorted β-bend with a $C_\alpha^{\,i} - C_\alpha^{\,i+3}$ distance of 8 Å, respectively.

DISCUSSION

Concanavalin A binds to certain polysaccharides to form precipitates (8–11), is capable of stimulating lymphocytes (15) and of agglutinating various somatic and germ-line cells (13–15), and has a receptor site specific for α-linked D-mannopyranoside. Its three-dimensional structure includes extensive regions of β-sheets (5, 6). Based on its amino-acid sequence, concanavalin A can also serve as a test case for the predictive power of the influence of nearest-neighboring amino acids on secondary structure of proteins (1–4).

Comparison of the experimentally observed β-sheets with the β-sheet-breaking residues taken from the 20 × 20 table of the influence of nearest-neighbor amino acids shows excellent agreement. In comparing the findings with the x-ray data of the Edelman group (5) and of Hardman and Ainsworth (6) of the 13 stretches involved in β-sheets (5), only five β-sheet-breaking residues (53, 127, 192, 193, and 194)

TABLE 2. *Comparison of the locations of β-sheet-breaking residues and experimentally observed β-sheets*

β-Sheet-breaking residues (*n*)	Frequencies of residues in β-sheet	β-Sheets experimentally observed: Edelman *et al.* (5)*	β-Sheets experimentally observed: Hardman and Ainsworth (6)*
	4 Ile (4,1,3)	4-9	4-10
	5 Val (3,4,2)		
	6 Ala (2,6,2)		
	7 Val (6,1,6)		
	8 Glu (1,1,3)		
	9 Leu (1,1,2)		
10 Asp (3,0,1)			
15 Thr (1,0,3)			
20 Pro (4,0,8)			
25 Ile (3,0,2)	25 Ile (3,0,2)	25-30*	24-30
	26 Gly (2,1,1)		
	27 Ile (5,2,2)		
	28 Asp (2,1,1)		
	29 Ile (1,1,1)		
	30 Lys (2,2,9)		
	36 Lys (3,3,2)	36-38*	35-38
	37 Thr (8,6,1)		
	38 Ala (1,2,8)		
42 Met (2,0,3)			
43 Gln (0,0,3)			
	48 Gly (4,4,2)	48-55	45-56
	49 Thr (4,5,7)		
	50 Ala (0,1,3)		
	51 His (3,3,2)		
	52 Ile (1,1,0)		
53 Ile (2,0,1)	53 Ile (2,0,1)		
	54 Tyr (4,3,9)		
	55 Asn (4,1,3)		
	59 Lys (2,1,2)	59-66	59-66
60 Arg (11,0,6)	60 Arg (11,0,6)		
61 Leu (2,0,1)	61 Leu (2,0,1)		
	62 Ser (9,1,4)		
	63 Ala (6,1,8)		
	64 Val (12,2,6)		
	65 Val (6,1,5)		
	66 Ser (1,1,3)		
72 Asx $(\frac{4,1,0}{3,0,4})$			
73 Ala $(\frac{2,1,7}{1,0,4})$	73 Ala $(\frac{2,1,7}{1,0,4})$	73-78	74-79
	74 Thr (4,3,9)		
	75 Ser (4,1,3)		
	76 Val (1,1,13)		
	77 Ser (1,1,3)		
	78 Tyr $(\frac{4,2,4}{4,1,4})$		
79 Asx (2,0,3)	79 Asx (2,0,3)		
83 Asx $(\frac{3,3,4}{7,0,9})$			
85 Val $(\frac{5,1,3}{2,0,2})$			
86 Leu (0,0,7)			
92 Val (0,0,3)	92 Val (0,0,3)	92-97	90-98
	93 Gly (1,1,3)		
	94 Leu (2,4,15)		
	95 Ser (9,1,4)		
	96 Ala (1,1,13)		
	97 Ser (3,1,3)		
102 Lys (2,0,1)			
106 Thr (2,0,2)	106 Thr (2,0,2)	106-116	106-116
	107 Leu (4,1,3)		
	108 Ile (6,2,7)		
	109 Ser (0,0,0)		
	110 Phe (1,1,13)		
	111 Ser (0,0,0)		
	112 Trp (6,4,3)		
	113 Thr (0,0,1)		
	114 Ser (1,2,8)		
	115 Lys (7,2,4)		
	116 Leu (6,1,6)		
120 Ser $(\frac{2,1,7}{1,0,4})$			
123 Glx (2,0,2)			
126 Ala $(\frac{5,1,3}{2,0,2})$	126 Ala $(\frac{5,1,3}{2,0,2})$	125-132	126-132
127 Leu (4,0,3)	127 Leu (4,0,3)		
	128 His (1,1,1)		
	129 Phe (0,0,0)		
	130 Met (1,1,1)		
	131 Phe (2,0,0)		
	132 Asn (1,1,1)		
135 Ser (5,0,4)			
140 Asp (11,0,6)	140 Asp 11,0,6)	140-144	140-146
	141 Leu (2,1,0)		
	142 Ile (8,3,2)		
	143 Leu (1,1,1)		
	144 Gln (4,2,7)		
153 Gly (1,0,2)			
	154 Asn (5,2,4)		154-156
	155 Leu (2,1,0)		
	156 Glu (8,3,1)		
157 Leu (2,0,1)			
158 Thr (7,0,2)			
160 Val (2,0,1)			
163 Asx $(\frac{3,0,2}{2,2,3})$			
164 Glx $(\frac{4,0,8}{0,0,4})$			
165 Ser $(\frac{1,0,2}{1,1,1})$			
	173 Arg (4,5,7)	173-177	171-177
174 Ala (3,0,4)	174 Ala (3,0,4)		
	175 Leu (3,1,2)		
	176 Phe (0,2,3)		
177 Tyr (2,0,4)	177 Tyr (2,0,4)		
178 Ala (0,0,7)			
185 Ser (3,0,1)			
187 Ala $(\frac{3,0,2}{2,2,3})$			
188 Glx (15,0,5)			
189 Ala $(\frac{3,0,1}{2,3,2})$			
	190 Ser (12,2,6)	190-199	191-199
	191 Val (1,0,1)		
192 Phe (1,0,2)	192 Phe (1,0,2)		
193 Glu (2,0,4)	193 Glu (2,0,4)		
194 Ala (2,0,1)	194 Ala (2,0,1)		
	195 Thr (3,1,2)		
	196 Phe (2,1,5)		
	197 Thr (2,2,2)		
	198 Leu (4,1,3)		
	199 Val (2,3,4)		
204 Asp (1,0,6)			
207 Pro (3,0,0)			
	209 Asp (8,3,9)	209-215	209-216
	210 Gly (2,1,0)		
	211 Ile (4,5,7)		
	212 Ala (0,0,0)		
	213 Phe (3,1,2)		
	214 Phe (0,0,1)		
	215 Ile (6,2,0)		
218 Ile $(\frac{1,0,3}{4,1,2})$			
220 Ser $(\frac{4,0,8}{0,0,4})$			
230 Leu (3,0,4)			
235 Pro (2,0,3)			
237 Ala (1,0,2)			

* Edelman *et al.*, private communication.

† Revisions based on corrections to fit sequence of Edelman *et al.* (5).

were inconsistent with the experimentally located β-sheets. Residue 53 was the sixth, 127 the second, and 192, 193, and 194 were the third, fourth, and fifth residues from the beginning of the sheet in the Edelman list (5). The same residues were inconsistent in the Hardman and Ainsworth list (6) plus residue 25. The agreement of the crystallographically determined β-sheets and the permissively β-sheet regions indicate the utility of examining sequences of proteins using the 20 × 20 table to locate approximately the permissively β-sheet and permissively helical regions in unknown proteins.

The increase in the amount of data to 12 known proteins, as compared with 5 proteins (1) and 11 proteins (4), improved the 20 × 20 table for locating permissively α-helical and β-sheet regions substantially, and as data on more proteins accumulate the table should become even more discriminating. It will clearly be substantially improved when the (ϕ, ψ) angles on concanavalin become available and are added.

The authors are indebted to Drs. Gerald Edelman, George N. Reeke, Jr., and Joseph W. Becker for pointing out the β sheets on their model of concanavalin A and to Dr. Karl Hardman for sending his data revised to conform to the sequence of Edelman *et al.* (5). Aided by grants from the National Science Foundation NSF GB-35243X to E.A.K. and NSF GB-31146 to T.T.W. and in part by a General Research Support Grant from the United States Public Health Service to Columbia University.

1. Wu, T. T. & Kabat, E. A. (1971) *Proc. Nat. Acad. Sci. USA* **68,** 1501–1506.
2. Kabat, E. A. & Wu, T. T. (1972) *Proc. Nat. Acad. Sci. USA* **69,** 960–964.
3. Wu, T. T. & Kabat, E. A. (1973) *J. Mol. Biol.*, **75,** 13–31.
4. Kabat, E. A. & Wu, T. T. (1973) *Biopolymers*, **12,** 751–774.
5. Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J. & Wang, J. L. (1972) *Proc. Nat. Acad. Sci. USA* **69,** 2580–2584.
6. Hardman, K. D. & Ainsworth, C. F. (1972) *Biochemistry* **11,** 4910–4919.
7. Goldstein, I. J., Hollerman, C. E. & Smith, E. E. (1965) *Biochemistry* **4,** 876–883.
8. Sumner, J. B. & Howell, S. F. (1936) *J. Bacteriol.*, **32,** 227–237.
9. Cifonelli, J. A. & Smith, F. (1957) *J. Amer. Chem. Soc.* **79,** 5055–5057.
10. So, L. & Goldstein, I. J. (1965) *Arch. Biochem. Biophys.* **111,** 407–414.
11. Lloyd, K. O., Kabat, E. A. & Beychok, S. (1969) *J. Immunol.* **102,** 1354–1362.
12. Edelman, G. M. & Millette, C. F. (1971) *Proc. Nat. Acad. Sci. USA* **68,** 2436–2440.
13. Inbar, M. & Sachs, L. (1969) *Proc. Nat. Acad. Sci. USA* **63,** 1418–1425.
14. Burger, M. M. & Noonan, K. D. (1970) *Nature* **228,** 512–515.
15. Powell, A. E. & Leon, M. A. (1970) *Exp. Cell Res.* **62,** 315–325.
16. Watson, H. C. (1969) *Progr. Stereochem.* **4,** 299–333.
17. Phillips, D. C. (1967) *Proc. Nat. Acad. Sci. USA* **57,** 484–495.
18. Blake, C. C. F., Mair, A. G., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967) *Proc. Roy. Soc. Ser. B*, **167,** 365–377.
19. Birktoft, J. J., Matthews, B. W. & Blow, D. M. (1969) *Biochem. Biophys. Res. Commun.* **36,** 131–137.
20. Lipscomb, W. N., Reeke, G. N., Jr., Hartsuck, J. A., Quiocho, F. A. & Bethge, P. H. (1970) *Phil. Trans. Roy. Soc. London Ser. B.* **257,** 177–214.
21. Wyckoff, A. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. & Richards, F. M. (1970) *J. Biol. Chem.* **245,** 305–328.
22. Wright, C. S., Alden, R. A. & Kraut, J. (1969) *Nature* **221,** 235–242.
23. Alden, R. A., Kraut, J. & Wright, C. S. (1970) *Phil. Trans. Roy. Soc. London Ser. B.* **257,** 119–124.
24. Hendrickson, W. A. & Love, W. E. (1971) *Nature New Biol.* **232,** 197–203.
25. Arnone, A., Bier, C. J., Cotton, F. A., Day, V. W., Hazen, E. E., Richardson, D. C., Richardson, J. S. & Yonath, A. (1971) *J. Biol. Chem.* **246,** 2302–2316.
26. Blundell, T. L., Cutfield, J., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Mercola, D. A. & Vijayan, M. (1971) *Nature* **231,** 506–511.
27. Blundell, T. L., Dodson, G. G., Dodson, E., Hodgkin, D. C. & Vijayan, M. (1971) *Recent Progr. Horm. Res.* **27,** 1–40.
28. Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A. & Margoliash, E. (1971) *J. Biol. Chem.* **246,** 1511–1535
29. Dickerson, R. E. (1972) *Sci. Amer.* **226,** no. 4, 58–72.
30. Mathews, F. S., Levine, M. & Argos, P. (1971) *Nature New Biol.* **233,** 15–16.
31. Mathews, F. S., Levine, M. & Argos, P. (1972) *J. Mol. Biol.* **64,** 449–464.
32. Birktoft, J. J. & Blow, D. M. (1972) *J. Mol. Biol.* **68,** 187–240.
33. Schiffer, M. & Edmundson, A. B. (1967) *Biophys. J.* **7,** 121–135.
34. Lewis, P. N., Momany, F. A. & Scheraga, H. A. (1971) *Proc. Nat. Acad. Sci. USA* **68,** 2293–2297.
35. Bunting, J. R., Athey, T. W. & Cathou, R. E. (1972) *Biochim. Biophys. Acta* **285,** 60–71.