J. Biosci., Vol. 9, Numbers 3 & 4, December 1985, pp. 203–212. © Printed in India.

Prediction and comparison of the secondary structure of legume lectins

M. JOGINADHA SWAMY, M. VENKATA KRISHNA SASTRY and AVADHESHA SUROLIA

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

MS received 3 May 1985; revised 29 September 1985

Abstract. Secondary structure prediction for the 4 legume lectins: Concanavalin A, soybean agglutinin, favabean lectin and lentil lectin, was done by the method of Chou and Fasman. This prediction shows that these four lectins fall into a structurally distinct class of proteins, containing high amounts of *β*-sheet and *β*-turns. There is a notable similarity in the gross structure of these proteins; all four of them contain about 40–50% of *β*-sheet, 35–45 % *β*-turn and 0–10% of α-helix. When the secondary structure of corresponding residues in each pair of these lectins was compared, there was a striking similarity in the Concanavalin A-soybean agglutinin and favabean lectin-lentil lectin pairs, and considerably less similarity in the other pairs, suggesting that these legume lectins have probably evolved in a divergent manner from a common ancestor. A comparison of the predicted potential *β*-turn sites also supports the hypothesis of divergent evolution in this class of lectins.

Keywords. Secondary structure prediction; legume lectins; divergent evolution; *β*-turns.

Introduction

Lectins are multivalent, carbohydrate binding and cell agglutinating proteins of plant and animal origin (Lis and Sharon, 1973; Goldstein and Hayes, 1978; Barondes, 1981). They elicit a variety of interesting cellular phenomena which include mitogenicity and preferential agglutination of cancerous cells (Rapin and Burger, 1974). In addition, lectins have been implicated in a variety of important functions in their parent organisms (Barondes, 1981).

The legume seeds have been found to be a rich source of lectins and for this reason the legume lectins have been the most widely studied. The legume lectins have also elicited a lot of interest due to the suggested evolutionary linkage between them (Hemperly and Cunningham, 1983). There are regions of extensive homology in the primary structure of these lectins (Hemperly and Cunningham, 1983; Foriers *et al.,* 1981). Secondary structure determination by circular dichroism (CD) reveals a considerable similarity in the sense that they all contain a large proportion of *β*-pleated sheet and negligible amount of α-helix (McCubbin *et al.,* 1971; Thomas *et al.,* 1979; Jirgensons, 1978). However, a detailed comparison of the secondary structure of these lectins, hitherto has not been done, though in one report the *β*-sheet regions of lentil lectin (LcL) were compared with those of Concanavalin A (Con A) (Foriers *et al.,* 1981).

Primary structure for 4 of the legume lectins *viz.,* Con A, soybean agglutinin (SBA),

Abbreviations used: LcL, Lentil lectin; Con A, Concanavalin A; CD, circular dichroism; SBA, soybean agglutinin; Favin, favabean lectin.

favabean lectin (Favin) and LcL has been determined. In this paper we report the prediction of secondary structure and its comparison for these 4 lectins.

It may be noted that the secondary structure of Con A and LcL was predicted earlier (Chou and Fasman, 1974a; Foriers *et al.,* 1981). However, it became necessary for us to predict their secondary structure again since the above predictions were done with propensity values obtained from a smaller data base of protein structures, whereas in the present study propensities from a larger data base of protein structures is used. Another reason is that the prediction of β-turns in the case of Con A was not done in conjunction with other secondary structural types while the β-turns in LcL were not predicted at all.

Methods

Primary structures of Con A, LcL and Favin have been taken from the work of Edelman *et al.* (1972), Foriers *et al.* (1981) and Hopp *et al.* (1982). Secondary structure from X-ray crystal structure data for Con A has been obtained from Reeke *et al.* (1975). Amino acid sequence of SBA was obtained from the gene sequence data of Vodkin *et al.* (1983). Secondary structure prediction has been done by the method of Chou and Fasman (1978) using the propensity values P_α and P_β for the 20 amino acids derived from the crystal structure data of 29 proteins. β-Turns were predicted as reported by Chou and Fasman (1979a) using the tetrapeptide bend probability which is given by the expression,

$$P_t = f_i \times f_{i+1} \times f_{i+2} \times f_{i+3},$$

where $f_i, f_{i+1}, f_{i+2},$ and f_{i+3} are the propensities of residues 1, 2, 3 and 4 of the tetrapeptide to be in position 1, 2, 3 and 4 of a β-turn.A cut off value of $P_t = 0{\cdot}75 \times 10^{-4}$ as recommended by Chou and Fasman (1978) was used. For all the tetrapeptides with $P_t > 0{\cdot}75 \times 10^{-4}$, the average propensities of α-helix ($\langle P_\alpha \rangle$), β-sheet ($\langle P_\beta \rangle$) and β-turn ($\langle P_t \rangle$) have been calculated. Tetrapeptides obeying the conditions $\langle P_\alpha \rangle < \langle P_t \rangle > \langle P_\beta \rangle$, and $P_t > 1{\cdot}00$ were assigned as β-turns. However, some of these turns were eliminated from the secondary structure assignment by considering them in conjunction with α-helical and β-sheet regions.

All the calculations in the present study were done on a Sord microcomputer (Sord Computer Systems Inc., Tokyo, Japan) using a programme* written in Basic.

For the sake of comparison the amino acid sequences of these lectins have been arranged as shown in figure 1 which is essentially the same as that reported earlier (Foriers *et al.,* 1981; Hemperly and Cunningham, 1983) excepting that no gaps were introduced in our comparison.

Residue to residue correlation analysis of the secondary structure has been carried out in the following way. The protein sequences have been aligned as shown in figure 1 and the secondary structure (predicted) of corresponding residues was checked for identity.

* The programme listing is available with the authors on request.

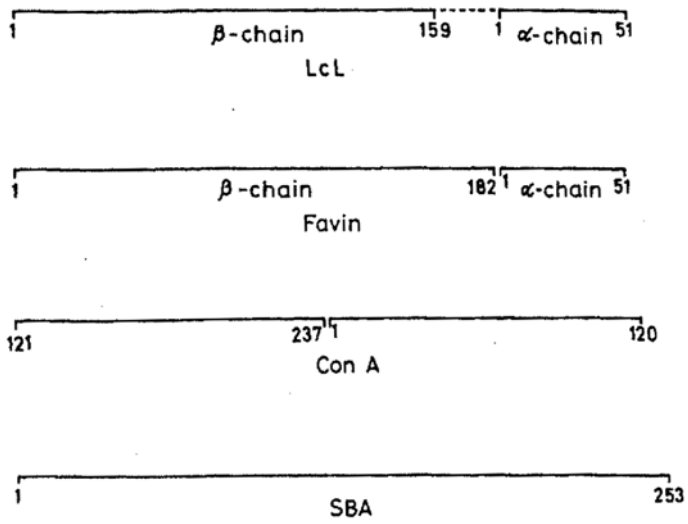

Figure 1. Alignment scheme for the 4 legume lectins; Con A, SBA, LcL and Favin, followed in the present study.

Results

The percentage content of the 3 secondary structural types (α -helices, β-sheets and β turns) in the 4 legume lectins considered in the present study, *viz.* Con A, SBA, Favin and LcL, as predicted by the method of Chou and Fasman (1978) is given in table 1. For the sake of comparison, these structural details obtained from CD and X-ray crystallographic investigations are also given wherever available.

Details and comparative analysis of the various predicted α-helical and β-sheet regions of Con A, SBA, Favin and LcL, and these structural details obtained from X-ray crystallographic studies on Con A are given in table 2.

Table 1. Comparison of α-helix, β-sheet and β-turn content in legume lectins by different methods.

	Sheet (%)		Helix (%)		Turn (%)
Protein	Prediction	CD	Prediction	CD	Prediction
Con A	40·1 (38)	30	10·5 (2)	0	43 (39·2)
SBA	41·4	26	9·8	0	39·5
Favin β	45·6		7·1	0	37·9
α	56·9	—	13·7	0	19·6
LcL β	49·0		0	0	43·4
α	45·0	—	15·7	0	31·4

Values in parentheses indicate the results obtained from X-ray investigations.

Table 2. Comparison between secondary structural regions of legume lectins.

Con A				
X-ray	Predicted	SBA	Favin	LcL
	122–130*			
124–130				
		3–12	4–9	3–11
			18–22	15–20
139–143				
		23–28		
147–149				
			27–32*	
153–156				
	156–160	32–37	32–42	31–36
				43–48
		50–55	44–54	50–53
169–176,	172–181	57–61		
179–181				
				60–65
	182–191*	67–73*		
			64–72	68–72
188–200	192–199	74–80		
			84–89	84–91
208–215	210–216			
		90–95*		
			101–106	101–106
	228–233	107–112		
		117–121	113–121	115–120
4–11	3–9			122–128
		122–127*		
			136–140	137–141
23–30	25–32	141–149	144–148	146–150
		152–157*		
35–40				
			159–165*	
48–56	47–56	160–168		
			169–174	
59–66	60–67	174–181	177–181	
			1–6	1–10
73–79	72–79	189–196	9–16	12–16
81–85*	80–85*			
87–97	88–93	203–208	19–25	19–26
			30–36*	29–36*
103–106				
108–116	103–115	223–228	37–44	
		244–250		

Regions indicated with an asterisk (*) are α-helical; rest all are β-sheets. This table gives only α-helical and β-sheet regions in these lectins.

The occurrence of β-turns in these proteins as predicted by the method of Chou and Fasman (1978) is given in table 3. There are 38 turns in Con A, 33 turns in SBA, 26 turns in Favin and 27 turns in LcL. The β-turns of Con A, as obtained from crystal structure data (Reeke *et al.,* 1975), are also given in table 3.

Table 3. Predicted β-turn sites in legume lectins.

Lectin	β-Turn sites
Con A*	10–13, 14–17, 15–18, 28–31, 31–34, 34–37, 43–46, 55–58, 56–59, 67–70, 86–89, 97–100, 117–120, 134–137, 137–140, 143–146, 147–150, 149–152, 150–153, 160–163, 166–169, 183–186, 203–206, 216–219, 222–225, 226–229, 227–230, 229–232.
Con A	10–13, 12–15, 14–17, 16–19, 19–22, 34–37, 42–45, 56–59, 67–70, 71–74*, 75–78*, 94–97, 96–99, 102–105*, 116–119, 117–120, 119–122, 131–134, 134–137, 136–139, 142–145, 147–150, 149–152, 151–154, 160–163, 161–164, 164–167, 167–170, 169–172, 201–204, 203–206, 206–209, 216–219, 218–221, 222–225, 224–227, 225–228, 233–236.
LcL α-chain	26–29, 37–40, 40–43, 47–50
β-chain	4–7*, 12–15, 21–24, 22–25, 25–28, 27–30, 39–42, 54–57, 60–63*, 64–67, 66–69*, 74–77, 80–83, 92–95, 96–99, 97–100, 106–109, 108–111, 112–115, 131–134, 142–145, 151–154, 156–159
Favin α-chain	27–30, 46–49, 48–51
β-chain	5–8*, 13–16, 16–19, 22–25, 23–26, 55–58, 57–60, 74–77, 76–79, 80–83, 92–95, 96–99, 97–100, 105–108, 107–110, 109–112, 122–125, 130–133, 132–135, 141–144, 150–153, 155–158, 174–177.
SBA	5–8*, 15–18, 20–23, 28–31, 38–41, 42–45, 44–47, 46–49, 63–66, 80–83, 82–85, 86–89, 98–101, 104–107, 112–115, 114–117, 116–119*, 128–131, 130–133, 134–137, 136–139, 142–145*, 157–160, 169–172, 171–174, 181–184, 183–186, 185–188, 216–219, 229–232, 234–237, 236–239, 240–243.

β-Turn sites indicated with an asterisk (*) are eliminated in the assignment on comparison with propensities for helix and sheet in the neighbouring residues.
* β-Turns obtained from X-ray studies.

After aligning the 4 legume lectins as outlined in figure 1, secondary structure of corresponding residues in each pair, for all the 6 pairs was compared, and the results of such a comparison are presented in table 4. The correlations were calculated in percentages, with respect to the number of residues of the protein containing lesser number of residues in each pair. For example, in the Con A-Favin pair, there are 76 positions where the corresponding residues are in identical conformation, and the correlation is 32·6 % with respect to the number of residues in Favin. The same method is employed when residues of each structural type (namely, α-helix, β-sheet and β-turn) are compared.

Correlation analysis of β-turns is depicted schematically in figure 2. All the β-turns predicted using a cutoff value of $0{\cdot}75 \times 10^{-4}$ have been represented here. Two higher cutoff values of $1{\cdot}25 \times 10^{-4}$ and $2{\cdot}0 \times 10^{-4}$ have also been chosen for comparison purposes. The β-turns in two different proteins are considered to be corresponding (conserved) if they are in a range of ± 2 residues. This analysis is presented in table 5. At

Table 4. Residue to residue correlation analysis of secondary structures of legume lectins.

Lectin pair	No. of corresponding residues in identical secondary structures			
	α-Helix	β-Sheet	β-Turn	Total
SBA–Con A	5 (20·8 %)	51 (57·3 %)	65 (65·0 %)	121 (51·05 %)
SBA–Favin	—	42 (44·2 %)	30 (37·9 %)	72 (30·9 %)
SBA–LcL	—	40 (42·1 %)	39 (46·4 %)	79 (37·6 %)
Con A–Favin	—	39 (43·8 %)	37 (46·8 %)	76 (32·6 %)
Con A–LcL	—	29 (32·6 %)	36 (42·9 %)	65 (30·9 %)
Favin–LcL	6 (85·7 %)	67 (69·8 %)	56 (66·67 %)	129 (61·4 %)

Values given in parentheses refer to per cent correlation calculated with respect to the number of residues of the protein containing lesser number of residues in each pair.

the lowest cutoff value, the correlation for all combinations is fairly high, with the highest values being associated with Con A-SBA and Favin-LcL pairs. This trend continues for the cases where higher cutoff values have been applied.

Discussion

The secondary structure content of the legume lectins (given as percentage α-helix etc., table 1) shows that there is a striking similarity in the gross structure of these proteins. All these proteins contain about 40–50% β-sheet, 35–45 % *β*-turn and 0–10% α-helix, and thereby fall into a structurally distinct class of proteins. The *β*-sheet, *α*-helical and *β*-turncontent predicted here agrees well with X-ray structure determination of Con A by Reeke *et al.* (1975). So far, the structural studies on these lectins using CD, have only classified them as proteins containing a high amount of *β*-sheet and low amount of α-helix. This study shows that, in addition to these features, the legume lectins are also characterized by a high *β*-turn content.

Though the gross structure, as seen from table 1, is very similar for the 4 legume lectins considered in the present study, it is possible that the order in which the various secondary structural segments are arranged in each of them could be quite different, thereby leading to the possibility of each of them having a unique 3-dimensional structure of its own, which could be considerably different from that of the rest. In order to further examine the similarities in the arrangement of various secondary structural segments, the following method has been employed. The primary structures

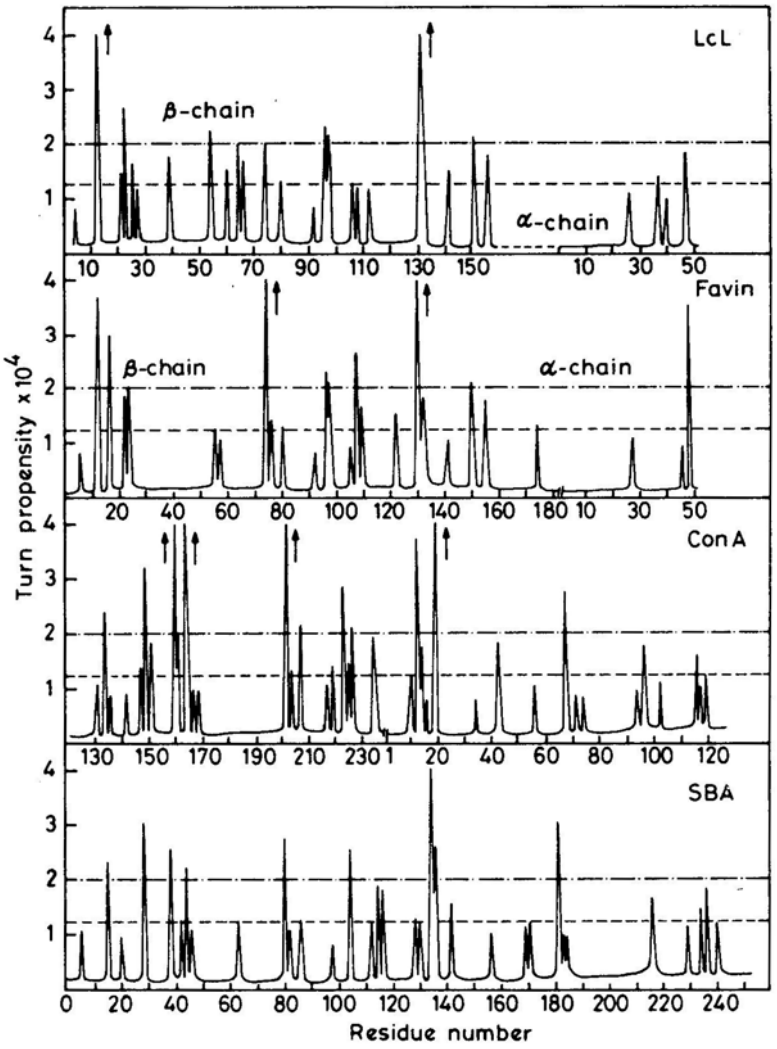

Figure 2. Propensities of β-turns in legume lectins. A cutoff value of $0{\cdot}75 \times 10^{-4}$ has been chosen to identify probable sites of β-turns which are indicated as sharp peaks. Two arbitrary cutoff values $1{\cdot}25 \times 10^{-4}$ (------) and $2{\cdot}0 \times 10^{-4}$ (-·-·-·-) have been chosen for analysis purposes. Wherever the propensity is higher than $4{\cdot}0 \times 10^{-4}$ the peak is drawn up to $4{\cdot}0 \times 10^{-4}$ and is marked with an arrow. Dotted line in the figure corresponding to LcL denotes the missing stretch of 23 residues corresponding to residues 160–182 of Favin.

of the legume lectins are aligned, and two regions of similar secondary structure from different proteins have been considered to be correlating with each other, if, atleast half of the residues of the smaller segment are overlapping with the residues of the other. Such an analysis for the α-helical and β-sheet regions of the 4 lectins studied here is represented in table 2. There are 13 regions of overlap between LcL and Favin, 8 between SBA and Favin, 7 between SBA and LcL, 10 between SBA and Con A, 4

Table 5. Correlation analysis of the conservation of turns in various lectins pairs.

	β-Turns conserved (%)		
	Cutoff values		
Lectin pairs	0.75×10^{-4}	1.25×10^{-4}	2.0×10^{-4}
Con A–SBA	69·7 (23)	61·1 (11)	77·8 (7)
Con A–LcL	48·2 (13)	31·1 (6)	28·5 (2)
Con A–Favin	46·2 (12)	27·8 (5)	28·6 (2)
SBA–LcL	48·2 (13)	16·7 (3)	0·0 (0)
SBA–Favin	42·3 (11)	22·2 (4)	14·2 (1)
Favin–LcL	65·3 (17)	66·7 (12)	71·4 (5)

Values in the parentheses indicate number of correlating sites.
These percentages are calculations with respect to the protein containing lesser number of β-turns in each pair.

between Con A and LcL and 6 between Con A and Favin. The two highest correlation values belong to the Con A – SBA and Favin-LcL pairs, suggesting a closer similarity between Con A and SBA, and between Favin and LcL than in other pairs.

In order to investigate whether the above conclusion drawn from the comparison of the different secondary structural segments is correct, we have carried out a residue to residue comparison of the secondary structure of each pair of the legume lectins. The results of this analysis are given in table 4. The correlation for the total sequences in the pairs Con AASBA and Favin-LcL is 51·1 % and 61·4 % respectively and is considerably higher than for all the other pairs, where the correlation ranges between 30·9 % and 37·6 %. Correlation of the residues in different secondary structural types also follows the same trend. When α-helical residues are compared, the correlation is 20·8 % and 85·7 % for the pairs Con A–SBA and Favin-LcL respectively, whereas for other pairs it is zero. The correlation of β-sheet residues in 57·3 % and 69·8 % in these two pairs respectively and less than 45 % in the other pairs. Correlation of β-turn residues also follows the same pattern, where the correlation is 65 % in Con AASBA pair and 66·7% in Favin-LcL pair and lies between 37 and 47 % in other pairs. These results suggest a closer relation between Con A and SBA, and between Favin and LcL than in the other pairs which can be taken to indicate a divergence in the evolution of these lectins.

Comparison of the predicted β-turn sites is depicted schematically in figure 2. When the β-turn sites were scanned to see correlating sites within ± 2 residues for each turn, in all the pairs, it was observed that there is a high correlation between the sites in Con A and those in SBA and between the sites in Favin and LcL. When the predicted β-turn sites of Con A were compared with those obtained from X-ray crystallographic investigations, a high degree of correlation (77 %) was observed. Earlier, a correlation analysis of the occurrence of β-turns for proinsulin-c-peptides, pancreatic ribonucleases and proteinase inhibitors by Chou and Fasman (1979b), showed that these chain reversal regions are highly conserved portions of proteins evolving from a common ancestor. Since, the legume lectins are also suggested to have evolved from a common

ancestor, we have carried out a correlation analysis of the β-turn regions of these proteins, and the details of such an analysis are given in table 5. When a cutoff value of $0{\cdot}75 \times 10_{-4}$ was applied for the β-turn propensity to identify the probable bend regions in these proteins, and the percentage correlation calculated, the highest correlation values were associated with the two pairs Con A–SBA and Favin-LcL, while the remaining pairs showed considerably less correlation.

The β-turn propensity of any tetrapeptide, W-W-X-Y-Z is obtained by multiplying the propensities of W, X, Y and Z to be in positions 1, 2, 3 and 4 respectively, of a β-turn. If, a residue in this tetrapeptide, say X, is replaced during evolution, by another residue with a lower propensity to be in the 2nd position of a β-turn, the overall turn propensity of the tetrapeptide will decrease. Such a replacement is more likely to bring down the turn propensity below the cutoffvalue in a tetrapeptide with low turn propensity, than in one with a high turn propensity. Therefore, the predicted sites with higher turn propensity are more likely to be conserved during the course of evolution than those with lower bend propensity. Therefore, if the cutoff value for identifying the β-turn regions is increased, the percentage correlation between two conserved proteins should increase; however, for non conserved proteins, where the correlation is due to random matching, this would result in a decreased correlation because the number of sites would be decreased. Hence we reasoned that if the higher correlation in the Con A–SBA and Favin- LcL pairs is due to a closer evolutionary relationship between the constituents of each pair, the percentage correlation should continue to remain high, and for other pairs it should decrease, which has indeed been observed. The correlation values for Con AASBA and Favin-LcL pairs, remain in the same range when the cutoff values were increased to $1{\cdot}25 \times 10^{-4}$ and $2{\cdot}0 \times 10^{-4}$, whereas for other combinations it decreases dramatically, further confirming that the evolutionary relations between Con A and SBA and between Favin and LcL are much closer than in the other combinations.

In summary, these results indicate that the 4 legume lectins fall into two classes, Con A and SBA forming one Favin and LcL forming the other. This bifurcation points to a divergence in the course of their evolution from a common ancestor. Since there are only a limited number of legume lectin sequences available at present, it is difficult to predict whether this divergence has branched into more than two classes, which could be better understood when the primary structures of more legume lectins are known.

Acknowledgements

This work was supported by a grant from the Department of Science and Technology, and partially (through a fellowship to M.V.K.S.) by the Indian National Science Academy, New Delhi. Authors thank Dr. M. R. N. Murthy for useful discussions. M.J.S. is a Fellow of the University Grants Commission, New Delhi.

References

Barondes, S. H. (1981) *Ann. Rev. Biochem.,* **50,** 207.
Chou, P. Y. and Fasman, G. D. (1974a) *Biochemistry,* **13,** 215.

Chou, P. Y. and Fasman, G. D. (1974b) *Biochemistry,* **13,** 222.
Chou, P. Y. and Fasman, G. D. (1978) *Adv. Enzymol.,* **47,** 45.
Chou, P. Y. and Fasman, G. D. (1979a) *Biophys. J.,* **26,** 367.
Chou, P. Y. and Fasman, G. D. (1979b) *Biophys. J.,* **26,** 385.
Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J. and Wang, J. L. (1972) *Proc. Natl. Acad: Sci. USA,* **69,** 2580.
Foriers,A., Lebrun, E.,van Rapenbusch, H., deNeve, N. and Strosberg,A.D. (1981)*J.Biol. Chem.,* **256,** 5550.
Goldstein, I. J. and Hayes, C. E. (1978) *Adv. Carbohydr. Chem. Biochem.,* **35,** 127.
Hemperly, J. J. and Cunningham, B. A. (1983) *Trends Biochem. Sci.,* **8,** 100.
Hopp, T. P., Hemperly, J. J. and Cunningham, B. A. (1982) *J. Biol. Chem.,* **257,** 4473.
Jirgensons, B. (1978) *Biochim. Biophys. Acta,* **536,** 205.
Lis, H. and Sharon, N. (1973) *Ann. Rev. Biochem.,* **42,** 541.
McCubbin, W. D., Oikawa, K. and Kay, G. M. (1971) *Biochem. Biophys. Res. Commun.,* **43,** 666.
Rapin, A. and Burger, M. M. (1974) *Adv. Cancer Res.,* **20,** 1.
Reeke, G. N., Jr., Becker, J. W. and Edelman, B. M. (1975) *J. Biol. Chem.,* **250,** 1525.
Thomas, M. W., Rudzki, J. E., Walborg, E. F. and Jirgensons, B. (1979) in *Proceedings of the Symposium on Carbohydrate-Protein Interactions* (ed. I. J. Goldstein) ACS Symposium series No. 88, pp. 67-75.
Vodkin, L. O., Rhodes, P. R. and Goldberg, R. B. (1983) *Cell,* **34,** 1023.