Cloning and sequencing of winged bean (*Psophocarpus tetragonolobus*) basic agglutinin (WBA I): presence of second glycosylation site and its implications in quaternary structure

Vivek Sharma, V.R. Srinivas, Avadhesha Surolia*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Received 20 May 1996

Abstract **We report cloning of the DNA encoding winged bean basic agglutinin (WBA I). Using oligonucleotide primers corresponding to N- and C-termini of the mature lectin, the complete coding sequence for WBA I could be amplified from genomic DNA. DNA sequence determination by the chain termination method revealed the absence of any intervening sequences in the gene. The DNA deduced amino acid sequence of WBA I displayed some differences with its primary structure established previously by chemical means. Comparison of the sequence of WBA I with that of other legume lectins highlighted several interesting features, including the existence of the largest specificity determining loop which might account for its oligosaccharide-binding specificity and the presence of an additional *N*-glycosylation site. These data also throw some light on the relationship between the primary structure of the protein and its probable mode of dimerization.**

Key words: Primary structure; Legume lectin; PCR cloning; Sequence homology; Carbohydrate-binding site; N-linked glycosylation site

1. Introduction

Lectins from leguminous plants constitute a large family of homologous proteins displaying remarkable divergence in their carbohydrate specificity [1,2]. WBA I is a dimeric GalNAc specific lectin isolated from the seeds of winged bean (*Psophocarpus tetragonolobus*) [3–6]. It possesses a molecular weight of about 58 000 (subunit M_r 29 000), pI greater than 9.5 and agglutinating activity towards trypsinized type A and B human erythrocytes [4,6]. WBA I specifically recognizes blood group A-substance derived oligosaccharides such as the trisaccharide GalNAcα1-3Galβ1-4Glc [7,8]. Thermodynamic and kinetic studies suggested that WBA I has an extended binding site and its interactions with saccharides involve considerable reorientation of water molecules [6,8–10]. The WBA I–monosaccharide interactions, as revealed by titration calorimetric studies, are essentially enthalpically driven and the binding is a simple bimolecular reaction mediated by van der Waals interactions and hydrogen bonding [10]. In contrast to other legume lectins, WBA I shows several thermal unfolding domains by differential scanning calorimetery, suggesting that the secondary and/or tertiary structural elements in WBA I exhibit less interactions between the amino acid residues [10] perhaps due to glycosylation at the inter-subunit interface as in *Erythrina corallodendron* lectin (ECorL) [11].

The primary structure of WBA I recently established in our laboratory by chemical means was found to share considerable homology with other legume lectins [12]. It displayed maximum homology to ECorL, yet exhibited several differences such as those in residues constituting the putative glycosylation and carbohydrate-binding sites. In view of the importance of glycosylation and sequence variability in binding-site loops towards determining the quaternary association [11] and carbohydrate specificity [13], respectively, it was imperative to determine the sequence of DNA encoding WBA I. We therefore carried out amplification by polymerase chain reaction (PCR) and characterisation of the coding sequence of WBA I. High-resolution X-ray structural analysis of the lectin in complex with *N*-acetylgalactosamine is in progress [14]. These studies will enable us to carry out expression and site-directed mutagenesis of its combining site and hence serve to improve our understanding of the structural basis of protein-carbohydrate interactions.

2. Materials and methods

2.1. Materials

All the restriction enzymes, DNA modifying enzymes, were obtained from Amersham, New England Biolabs, Pharmacia and Stratagene. DNA sequencing reagents were obtained from US Biochemical Corp. Most of the other biochemicals and reagents were from Sigma.

2.2. PCR amplification and cloning of WBA I

Total genomic DNA was purified from germinating winged bean [15]. The WBA I coding sequence was amplified using 100 ng of the genomic DNA as the template and the primers 5'CTAGGAATTCATGAAAACCATATCGTTTAACTTC3' and 5'AGCCTCTAGAATTCGTTTGTTCCTGGCAACGA3'. The PCR conditions used were 95°C, 45 s; 55°C, 1 min; 72°C, 2 min, 30 cycles followed by 10 min extension at 72°C in a Pharmacia GENE-ATAQ thermal cycler. The PCR product was cloned into the pCRII TA-cloning vector from Invitrogen according to the manufacturer's instructions in order to obtain the clone pCRWBA-I. For sequencing and further characterisation, the fragment released with *Eco*RI was cloned into the same site of plasmid pT7T3(18U).

2.3. DNA sequencing

All sequencing was performed using Sequenase version 2.0 from US Biochemical Corp. according to the manufacturer's instructions. Entire sequencing was carried out on single-stranded DNA obtained from the clones in pT7T3(18U) obtained by infection with the helper phage M13K07 [15]. Both strands were sequenced using M13(-40) universal and M13 reverse primers. The overlapping sequence information was obtained from subcloning restriction fragments.

3. Results and discussion

3.1. Cloning and sequencing of WBA I gene

Oligo primers were designed corresponding to the N- and C-termini of the WBA I using the codon degeneracy and the

*Corresponding author. Fax: (91) (80) 3341683.
E-mail: surolia@mbu.iisc.ernet.in

PII S0014-5793(96)00613-8

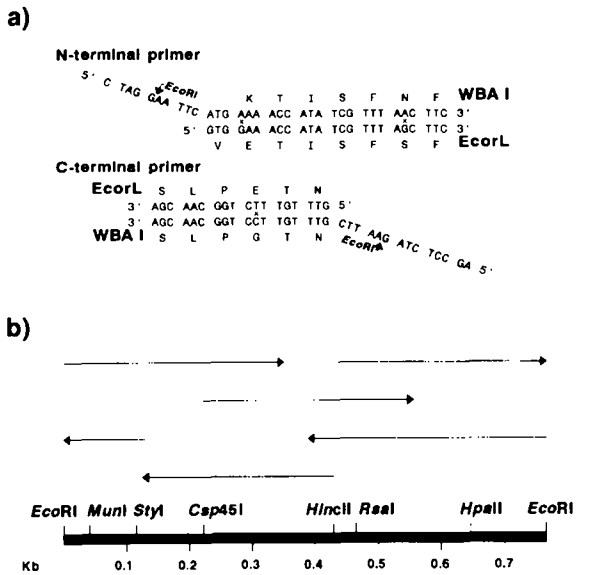

Fig. 1. (a) Design of oligonucleotide primers for WBA I. The DNA sequence of the corresponding regions of ECorL was modified by incorporating minimum changes (indicated by ×) for WBA I. (b) Sequencing strategy of *WBA I*. Unique restriction sites in *WBA I*, including sites for *Csp*45I and *Hin*cII which were used for the preparation of subclones necessary for sequencing, are indicated.

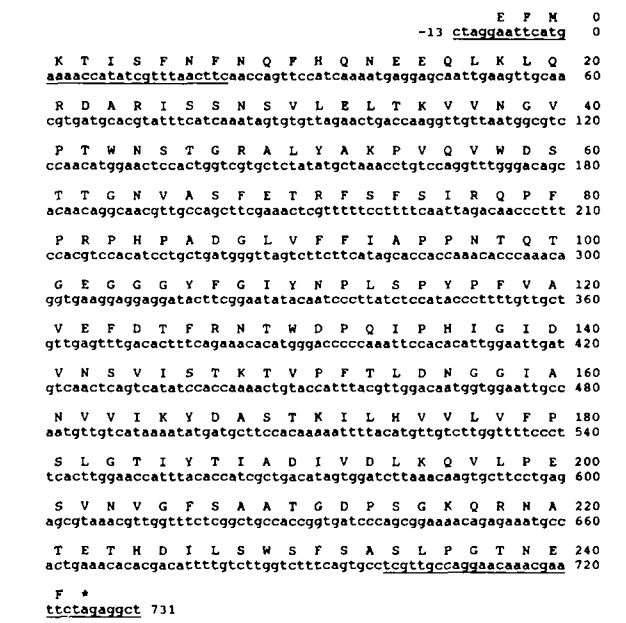

Fig. 2. Complete DNA and deduced amino acid sequence of WBA I. The sequence corresponding to the olgionucleotide primers used for PCR amplification is underlined.

sequence of homologous ECorL (Fig. 1a) [16]. The N-terminal 34-mer primer with the sequence 5′CTAGGAATTCATGAAAACCATATCGTTTAACTTC3′ consisted of 21 nucleotides (nt) corresponding to the first seven amino acids (KTISFNF) of the protein and an additional 16 nt anchor sequence containing sites for *Rca*I and *Eco*RI. The C-terminal 32-mer primer with the sequence 5′AGCCTCTAGAATTCGTTTGTTCCTGGCAACGA3′ containing 18 nt corresponding to the last six amino acids (SLPGTN) was designed with an additional 14 nt anchor sequence to incorporate *Xba*I and *Eco*RI sites. Since most of the legume lectin-encoding genes cloned thus far are known to lack intron sequences, the *WBA I* coding sequence was amplified using winged bean genomic DNA as the template. This gave a single PCR product of 744 bp, which being the size expected from the protein, indicated the absence of any intervening sequences in the *WBA I* gene. The PCR product was cloned into pCR II cloning vector (Invitrogen, USA) to obtain the clone pCRWBA-I.

The complete sequence of *WBA I* DNA was determined by sequencing both strands of the fragment released with *Eco*RI which was cloned into the same site of plasmid pT7T3(18U) from Pharmacia. The deletion clones obtained by cleaving at *Csp*45I and *Hin*cII restriction sites were used to obtain additional sequence information (Fig. 1b).

3.2. Comparison of DNA deduced protein sequence

WBA I encodes a 239 amino acid polypeptide with a calculated molecular mass of 26 178 Da (Fig. 2). WBA I was earlier proposed to be a 238 residue polypeptide of estimated molecular mass of 26 139 Da whereas a value of 29 000 Da has been approximated by SDS-PAGE. The excess mass of about 2.8 kDa obtained by SDS-PAGE was previously shown to be accounted for by the presence of covalently linked carbohydrates. The amino acid composition calculated from the deduced DNA sequence of WBA I closely resembles the chemically determined composition (Table 1). The sequence deduced from *WBA I* is fairly similar to the sequence determined at the protein level (Fig. 3). However, there are about 26 amino acid differences observed at 15 stretches within the WBA I sequence. Most of these differences are single amino acid changes and could represent sequence microheterogeneity in WBA I isolectins. The seven conservative differences: Ser^{29} (Gly), Glu^{32} (Gln), Asn^{14} (Gln), Val^{57} (Ile), Asp^{190} (Gln), Ile^{191} (Leu) and Asp^{193} (Asn) could be the result of such variation amongst isolectins (where the corresponding residues in the protein sequence are given in parentheses). At the same time, it is likely that some of the other differences arise due to protein sequencing errors. Thus, the two longer stretches: Ser^{114}–Val^{121} and Lys^{195}–Pro^{199}, which differ more significantly in both sequence and size from the polypeptide sequence, correspond to the regions of relatively poorer peptide overlaps in the protein sequence. The protein sequence of both these stretches is largely derived from single peptides obtained by tryptic cleavage [12]. The protein sequence of the region corresponding to Ser^{114}–Val^{121} was obtained by a single trypsin-cleaved soluble peptide which it now appears would have originated from autolysis of trypsin.

Comparison of the deduced amino acid sequence of WBA I

```
KTISFNFNQFHQNEEQLKLQRDARISSNSVLELTKVVNGVPTWNSTGRAL  50
----------------------------G--Q---------K-Q------  50

YAKPVQVWDSTTGNVASFETRFSFSIRQPFPRPHPADGLVFFIAPPNTQT 100
------I-----------------------------------------L--I- 100

GEGGGYFGIYNPLSPYPFVAV...EFDTFRNTWDPQIPHIGIDVNSVIST 147
-P----HL---K-GEDNIFVVEGN-------------------------- 150

KTVPFTLDNGGIANVVIKYDASTKILHVVLVFPSLGTIYTIADIVDLKQV 197
-------------------------------------------QL-N-Q.. 198

LPESVNVGFSAATGDPSGKQRNATETHDILSWSFSASLPGTN 239
..------------------------------------------ 238
```

Fig. 3. Comparison of DNA-deduced amino acid sequence of WBA I with chemical sequence. Complete sequence of the DNA deduced sequence is given in the single-letter amino acid code in the top rows and the differences in chemical sequence are shown below at the appropriate positions. The identity in chemical sequence is represented by a dash while the gaps are given as dots.

with other legume lectins reveals a high degree of homology (Fig. 4). The lectin sequence shows highest identity (64.4%) with *E. corallodendron* lectin (ECorL) and 51.46, 50.2, 48.9 and 43.5% identities with *Phaseolus vulgaris* leucoagglutinin (PHA L), soybean agglutinin (SBA), *Ulex europeus* isolectin II (UEA II) and concanavalin A (Con A), respectively [16–20]. There are 54 residues of WBA I which are invariant in all legume lectins which include those involved in the binding to carbohydrate, metal ions and hydrophobic ligands (Fig. 4). The WBA I residues corresponding to the carbohydrate-binding site are Asp^{87}, Gly^{105}, Phe^{126} and Asn^{128}, whereas Glu^{122}, Asp^{124}, Phe^{126}, Asn^{128} and His^{130} are equivalent residues for Ca^{2+} and Mn^{2+} binding. Comparing the sequence with that of other Gal/GalNAc binding lectins suggests that Asp^{212} could be an additional residue involved in the monosaccharide recognition perhaps by mediating a hydrogen bond to O-4 of galactose through one of its carboxylate oxygens.

The chemical sequence of WBA I suggested a single glycosylation N-linked site close to the C-terminus [12]. This was in contrast to other legume lectins which show glycosylation close to the N-termini [16,21,22]. The DNA deduced sequence reveals an Asn (instead of Gln) at position 44 which is the second and, being closer to the N-terminus, a more likely glycosylation site of WBA I.

3.3. Structural implications of WBA I sequence

The diverse carbohydrate specificities of legume lectins have been linked to the structural differences at the variable edge of their binding pocket [11]. The hypervariability of one of the four combining site loops is known to influence, in an inexplicable way, the carbohydrate recognition in these proteins [13]. Interestingly, this loop in WBA I is longer than all other legume lectins due to insertion of a stretch of four amino acids, Pro-Ser-Gly-Lys, just after the residue Asp^{212} which corresponds to Ala^{218} of ECorL that makes a backbone hydrogen bond with O-4 of Gal in the crystal structure [11]. The equivalent region in the lectin IV of *Griffonia simplicifolia*, Val^{221}-Gly^{222}-Tyr^{223}, as a result of the *cis* peptide bond between residues 221 and 222 adopts a unique structure in this region and interacts with the α1-4 linked fucose in the GSIV-[Fucα1-2Galβ1-3(Fucα1-4)GlcNAc] complex [23]. In a similar fashion the Pro-Ser-Gly-Lys region of WBA I may constitute an extended binding site with a probable *cis* peptide between Pro^{213} and Ser^{214} providing a similar topology which accommodates α1-2 linked fucose of the A-pentasaccharide, GalNAcα1-3(Fucα1-2)Galβ1-4(Fucα1-3)Glc.

Table 1
Amino acid composition of winged bean basic lectin

Amino acid	Kortt [5]	Puri and Surolia [12]	Present study
Asp	32	11	12
Asn	N.D.	18	16
Glu	19	8	8
Gln	N.D.	12	10
Ser	20	20	21
His	5	5	5
Gly	20	17	17
Thr	20	21	21
Arg	9	8	8
Ala	15	13	14
Tyr	7	5	6
Met	0	0	0
Val	19	21	22
Phe	18	16	17
Ile	17	18	16
Leu	16	15	15
Lys	11	10	9
Pro	21	16	18
Trp	4	4	4
Total	253	238	239

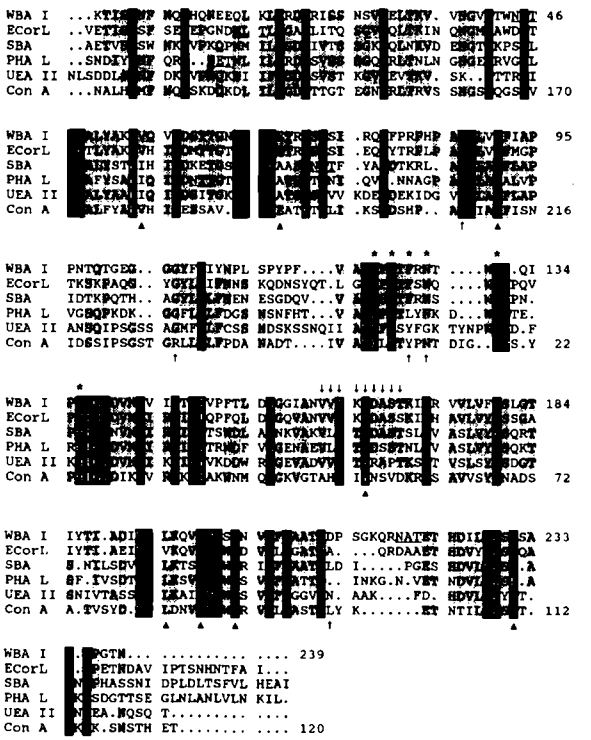

Fig. 4. Comparison of the amino acid sequence of WBA I with other legume lectins. WBA I is aligned with sequences of *Erythrina corallodendron* (ECorL), soybean (SBA), *Phaseolus vulgaris* leucoagglutinin (PHA L), *Ulex europeus* isolectin II (UEA II) and concanavalin A (Con A). Only the sequences of WBA I and Con A are numbered. The invariant amino acids are shown with black background whereas partially conserved residues are shaded. Residues involved in the metal binding in Con A are indicated by '*', those involved in putative adenine binding by '↓' and those involved in formation of hydrophobic cavity by '▲'. The glycosylation sites are underlined while the residues involved in carbohydrate binding are indicated by '↑'. Gaps (.) have been introduced for maximum alignment.

Despite possessing similar primary and tertiary structures the legume lectin protomers associate in different fashions [11,23 26]. Differential scanning calorimetric studies of the thermal denaturation of WBA I have revealed the existence of two distinct types of unfolding domains in this protein [10] as compared to other legume lectins such as concanavalin A which unfolds as a single entity [27], indicating striking differences in their mode of oligomerization. This is not surprising because galactose-binding legume lectins, in general, deviate from their mannose-binding counterparts in the quaternary arrangement of their subunits. These differences include the the lack of 12-stranded extended β-sheet arising due to either steric hindrance from the N-linked glycan as observed in ECorL [11] or the evasion of burying a charged residue as observed in GSIV [23,24]. Its high sequence homology with ECorL suggests that WBA I may display a similar quaternary structure. Moreover, ECorL displays unfolding behavior analogous to that of WBA I, indicating that these proteins might

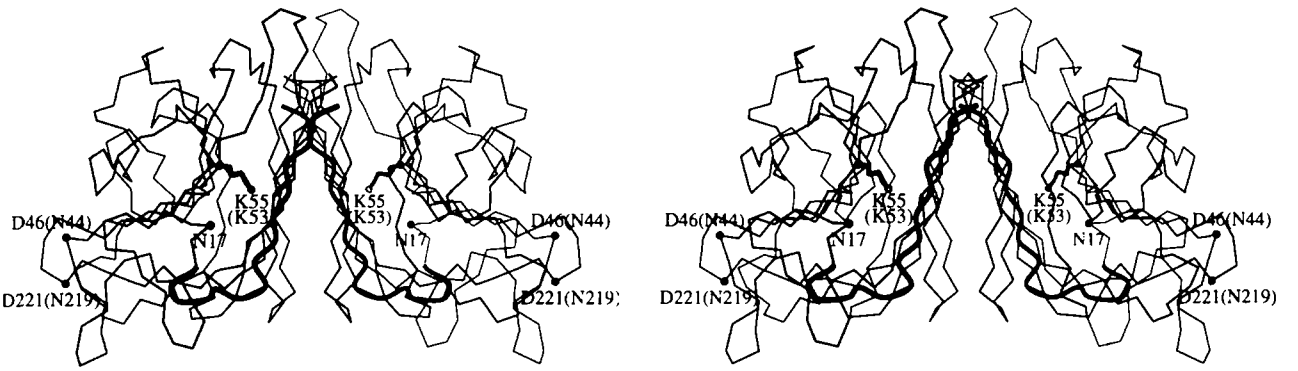

Fig. 5. Stereo view of the ECorL dimer. The thin lines represent the Cα-trace of ECorL whereas the thick coils correspond to the regions involved in dimerization in Con A. The position of some of the residues from ECorL and corresponding residues from WBA I (in parentheses) is indicated. Both the glycosylation sites of WBA I (N^{44} and N^{219}) are located far away from the region involved in Con A-type dimerization. In contrast, the glycosylation site of ECorL (N^{17}) as well as the lysine residue at position 55 (K^{53} in WBA I) are situated close to this region and can thereby influence mode of dimerization. This figure was generated using the program MOLSCRIPT [28].

have a similar mode of oligomerization (Surolia, A., Sharon, N. and Schwarz, F.P., unpublished observations). However, the two glycosylation sites in WBA I (Asn^{44} and Asn^{219}) are located at positions different from that observed for ECorL (Asn^{17}) (Fig. 5). Hence, glycosylation at these sites is unlikely to sterically hinder the formation of an extended β-sheet. The residue Lys^{53} of WBA I and Lys^{55} of ECorL are equivalent to Glu^{58} of GSIV, whereas it is Leu^{49}, Ser^{18} and Ala^{177} in favin, pea lectin and concanavalin A, respectively. The evasion of burying this lysine residue appears to be a more likely reason for the deviation from the Con A type of dimerization observed in both WBA I and ECorL (Fig. 5). Thus, the modes of oligomerization of legume lectins seem to be intrinsic to their protein sequence [26]. The determination of the three-dimensional structure of WBA I would provide a molecular explanation for the origin of this variation.

In summary, the primary structure of WBA I shows that it belongs to the family of single-chain legume lectins which is encoded by an intronless gene. The sequence of the WBA I gene provides additional information about several structural features of WBA I. These include amino acid differences related to its sugar specificity, a longer carbohydrate recognition loop probably constituting the extended binding site and the presence of an additional potential *N*-glycosylation site at Asn^{44}. Comparison of WBA I with ECorL and GSIV reveals that residue Lys^{53} may have implications in its modes of dimerization. Finally, the isolation of WBA I gene will facilitate the understanding of structure-function relationship by further studies aimed at its expression and mutagensis.

Acknowledgements: This work was been supported by grant from the Council for Scientific and Industrial Research, Government of India. The authors thank the Department of Biotechnology-supported oligonucleotide synthesis facility, Indian Institute of Science, for the oligonucleotides.

References

[1] Goldstein, I.J. and Poretz, R.D. (1986) in: The Lectins: Properties, Functions and Applications in Biology and Medicine (Liener, I.E., Sharon, N. and Goldstein, I.J. eds.) pp. 33–247, Academic Press, Orlando, FL.
[2] Sharon, N. and Lis, H. (1990) FASEB J. 4, 3198 3208.
[3] Pueppke, S. (1979) Biochim. Biophys. Acta 581, 63–70.
[4] Kortt, A.A. (1984) Eur. J. Biochem. 138, 519–525.
[5] Kortt, A.A. (1985) Arch. Biochem. Biophys. 236, 544 554.
[6] Khan, M.I., Sastry, M.V.K. and Surolia, A. (1986) J. Biol. Chem. 261, 3013–3019.
[7] Matsuda, T., Kabat, E.A. and Surolia, A. (1989) Mol. Immunol. 26, 189–195.
[8] Puri, K.D. and Surolia, A. (1994) Pure. Appl. Chem. 66, 497–502.
[9] Puri, K.D., Khan, M.I., Gupta, D. and Surolia, A. (1993) J. Biol. Chem. 268, 16378–16387.
[10] Schwarz, F.P., Puri, K.D. and Surolia, A. (1991) J. Biol. Chem. 266, 24344 24350.
[11] Shaanan, B., Lis, H., Sharon, N. (1991) Science 254, 862 866.
[12] Puri, K.D. and Surolia, A. (1994) J. Biol. Chem. 269, 30917 30926.
[13] Young, N.M. and Oomen, R.P. (1992) J. Mol. Biol. 228, 924–934.
[14] Sankaranarayanan, R., Puri, K.D., Ganesh, V., Banerjee, R., Surolia, A. and Vijayan, M. (1993) J. Mol. Biol. 229, 558–560.
[15] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[16] Arango, R., Rozenblatt, S. and Sharon, N. (1990) FEBS Lett. 264, 109–111.
[17] Hoffman, L.M. and Donaldson, D.D. (1985) EMBO J. 4, 883 889.
[18] Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B. (1983) Cell 34, 1023 1031.
[19] Konami, Y., Yamamoto, K. and Osawa, T. (1991) Biol. Chem. Hoppe-Seyler, 372, 95–102.
[20] Yamauchi, D. and Minamikawa, T. (1990) FEBS Lett. 260, 127 130.
[21] Hardman, K.D., Agarwal, R.C. and Freiser, M.J. (1982) J. Mol. Biol. 157, 69–86.
[22] Tague, B.W., Drickinson, C.D. and Chrispeels, M.J. (1990) Plant Cell 2, 533 546.
[23] Delbaere, L.T.J., Vandonselaar, M., Prasad, L., Quail, J.W., Wilson, K.S. and Dauter, Z. (1993) J. Mol. Biol. 230, 950–965.
[24] Delbaere, L.T.J., Vandonselaar, M. and Prasad, L. (1990) Can. J. Chem. 68, 1116–1121.
[25] Rini, J.M. (1995) Annu. Rev. Biophys. Biomol. Struct. 24, 551 577.
[26] Banerjee, R., Mande, S.C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S.K., Suguna, K., Surolia, A. and Vijayan, M.V. (1994) Proc. Natl. Acad. Sci. USA 91, 227–231.
[27] Borrebaack, C. and Mattiasson, B. (1980) Eur. J. Biochem. 107, 67–71.
[28] Kraulis, P. (1991) J. Appl. Crystallogr. 24, 946 950.