Complete amino acid sequence of the A chain of mistletoe lectin I

Montserrat Huguet Soler[a], Stanka Stoeva[a], Cornelia Schwamborn[b], Sabine Wilhelm[b], Thomas Stiefel[b], Wolfgang Voelter[a,*]

[a]*Abteilung für Physikalische Biochemie des Physiologisch-chemischen Instituts der Universität Tübingen, Hoppe-Seyler-Str. 4, D-72076 Tübingen, Germany*
[b]*Medisculab Arzneimittel GmbH, Fellbach, Germany*

Received 20 October 1996

Abstract The complete amino acid sequence of the A chain of mistletoe lectin I was determined via Edman degradation sequencing of the N-terminus and tryptic and endoproteinase Asp-N overlapping fragments, amino acid analysis and MALDI-MS. The data obtained show a great homology with the chains of ribosome-inactivating proteins such as ricin and abrin with 111 (abrin-a) and 103 (ricin-D) amino acid residues conserved, respectively. The knowledge of the primary structure of MLA will have a fundamental impact on elucidating the biological function of medically applied mistletoe lectins on a molecular basis.

Key words: Mistletoe lectin I; Immunomodulation; Ribosome-inactivating protein; Amino acid sequence; *Viscum album*

1. Introduction

During the last years, research has focused on lectins as therapeutically active substances present in mistletoe extracts [1]. Three mistletoe lectins with different sugar specificities (ML-I, ML-II and ML-III) have been isolated from these plant extracts [2]. ML-I, the β-galactoside-specific lectin from *Viscum album*, is a ribosome-inactivating protein (type II RIP). It consists of two different subunits, both *N*-glycosylated: the toxic A chain (MLA) is linked by a disulfide bridge to the galactose-binding lectin B chain (MLB) [3]. MLB (34 kDa) binds the toxin to the cell membrane and delivers MLA (29 kDa) into the cell cytosol where it inhibits protein synthesis by the same mechanism as the ricin A chain and other related proteins [4,5]. MLA catalytically inactivates 60S ribosomal subunits by hydrolyzing the *N*-glycosidic bond of the adenosine residue 4324 of 28S rRNA. It has been proved that the galactose-binding function is also required for the translocation of the glycosidic subunit into the cytosol [6]. MLA itself is non-toxic to intact cells [1], because it has no means of entering eukaryotic cells in order to reach ribosomal substrates.

ML-I is a component of a commercially available mistletoe extract with immunostimulating potency applied for the treatment of human cancer [7]. This immunomodulating capacity has recently been attributed to the presence of ML-I in the extract [8]. Regular non-toxic subcutaneous injections of optimal doses of ML-I (1 ng/kg body weight, twice weekly) yield enhancements of cytokine secretion in serum, significant increases in NK cell number, in phagocytic activity of granulocytes and in the number of large granular lymphocytes [7–11]. These effects prove the capacity of ML-I to stimulate non-specific defence mechanisms. ML-I administration is also followed by a release of β-endorphin in plasma [12]. β-Endorphin is an oligopeptide with profound effects on the nervous system. Although its function has not been clarified, this and other peptides of the pro-opiomelanocortin family are probably involved in general life events like mood, sleep and pain [13]. They act as natural painkillers or opiates and decrease the pain responses in the central nervous system. Consequently, ML-I treatment may induce improvement of the quality of life of cancer patients.

Recently, interest in ML-I has largely increased because of its obvious ability to activate non-specific defence mechanisms. The general pathway(s) of activation of these mechanisms are still unknown, but specific mistletoe lectin-carbohydrate (cell glycoconjugate) interactions appear to be the effector system involved in eliciting such an immune response [14]. Controlled increases of these non-specific defence mechanisms may induce clinically beneficial immunomodulation in the treatment of cancer [7]. Accordingly, the detailed chemical structure of ML-I is crucial to elucidate its immunomodulatory capacity and for the design of more effective therapeutic substances. The present study reports the complete primary structure of its cytotoxic subunit MLA, as the first step for the structural characterization of ML-I.

2. Materials and methods

2.1. Protein isolation

Mistletoe extract was obtained from an aqueous suspension by mixing 1.0 parts of mistletoe ground material with 1.3 parts of water (w/v). 20 ml mistletoe extract were centrifuged at 2000×*g* for 20 min. The supernatant was purified by filtration (Sterivex GV Filter, Millipore) and adjusted to 100 ml with peptide eluent (0.05 M $K_2HPO_4 \cdot 3H_2O$ in 0.15 M NaCl, pH 7.0). The solution was applied to an affinity chromatography column on divinyl sulfone-activated lactose-coupled Sepharose 4B (3×50 cm) to isolate ML-I from the mistletoe extract [15]. The column had previously been equilibrated and was then eluted with the same buffer. The immobilized lectin on the gel matrix was washed with one column volume of 5% β-mercaptoethanol (v/v) and incubated at room temperature overnight, whereby disulfide bridges linking the two chains were cleaved. After reduction, the column was equilibrated again with peptide eluent. Column eluate contains free MLA, while MLB remains bound on the gel material. The MLB can be obtained by elution with 0.2 M lactose solution [3]. The fractions containing protein were pooled, dialyzed against deionized water and lyophilized.

2.2. Amino acid analysis

Amino acid composition of MLA was determined with an Eppen-

*Corresponding author. Fax: (49) (7071) 293361.

Abbreviations: ML-I, mistletoe lectin I; MLA, A chain of mistletoe lectin I; MLB, B chain of mistletoe lectin I; MALDI-MS, matrix-assisted laser desorption ionization mass spectrometry; RIP, ribosome-inactivating protein

PII S0014-5793(96)01309-9

dorf/Biotronik LC 3000 amino acid analyzer after hydrolysis in 6 M HCl for 24 h at 110°C.

2.3. Enzymatic digestion

Digestion of MLA (4.40 mg) with endoproteinase Asp-N (sequencing grade, Boehringer Mannheim, Germany) was carried out in 50 mM sodium phosphate (pH 8.0) containing 2 M urea at 37°C for 24 h. Tryptic (sequencing grade, Boehringer Mannheim) digestion (7.60 mg) was performed in 100 mM Tris-HCl buffer (pH 8.5) with up to 1 M urea and 10% (v/v) acetonitrile for solubilization at 37°C for 18 h. After enzymatic digestion, a core fraction could be precipitated at pH 4.8 by titration with 0.1 M HCl and removed from the corresponding reaction mixture by centrifugation.

2.4. Fractionation of the peptides

The acidified peptides were fractionated by RP-HPLC on a Nucleosil C_{18} column (4 mm×250 mm) and eluted by a linear gradient of acetonitrile containing 0.15% (v/v) TFA at a flow rate of 1 ml/min. Fractions obtained after titration were resolubilized in 50% (v/v) formic acid and separated by RP-HPLC on a Vydac C_4 column (2 mm×250 mm). The peptides were eluted at a flow rate of 300 μl/min using the same buffer system. The UV absorbance of the peptides was monitored at 214 nm. All separations were done with an Eppendorf quaternary HPLC system of the series BT 9000.

2.5. Sequence analysis

HPLC-purified peptides (50–200 pmol) were dissolved in 0.1% (v/v) TFA and spotted onto polybrene-coated filters. The amino acid sequences of the peptides were determined on an automated sequencer (model 473A, Applied Biosystems).

2.6. Mass spectrometric analysis

Mass spectra were obtained by matrix-assisted laser desorption ionization (MALDI) mass spectrometry (Kratos, MALDI III equipment, Shimadzu, Europe). Peptides (10–50 pmol) were dissolved in 0.1% (v/v) TFA and applied on a target. Analysis was carried out in α-cyano-4-hydroxycinnamic acid matrix solution.

3. Results

The complete amino acid sequence of the MLA was determined by analysis of two sets of peptides generated by digestion of the protein with trypsin and endoproteinase Asp-N, respectively (Fig. 1). After digestion, reaction mixtures were fractionated by RP-HPLC and analyzed by sequence and mass spectrometric analysis. The MLA contains 254 amino acid residues. The molecular weight calculated on the basis of sequence data (28 480 Da) compares well with the apparent molecular weight determined by SDS-PAGE (29 000 Da) [16]. Similarly, the amino acid composition deduced from the sequence analysis coincides well with that determined by acid hydrolysis (Table 1).

The primary structure of MLA could be obtained mainly by analysis of the tryptic peptides. After hydrolysis, a core fraction could be precipitated by titration with HCl at pH 4.8. It contained two highly hydrophobic fragments (T8 and T20+T21) and other large peptides resulting from the incomplete cleavage of the protein. These latter ones allowed the alignment of other shorter peptides. The first tryptic fragments could be positioned with the help of the N-terminal sequence of the intact MLA, which led to an unambiguous identification of residues 1–28. Specific bond cleavages after residues Arg^{154} and Arg^{234} were not detected. All peptides were sequenced to the end. Determined amino acid sequences were confirmed by mass spectrometric analysis.

Sequence analysis of fragment T11 predicted the existence of a potential *N*-glycosylation site by the seventh Edman degradation cycle, which could be confirmed by MALDI-MS

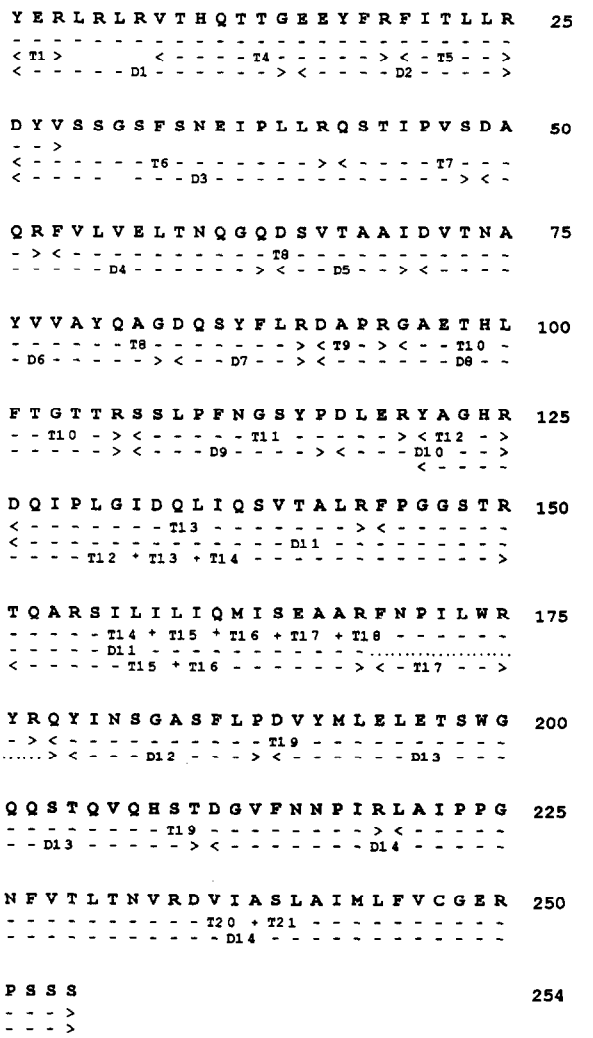

Fig. 1. Amino acid sequence of the MLA and sequencing strategy. T and D indicate peptides analyzed from trypsin and Asp-N digestions, respectively. Dots indicate unidentified residues. Only peptides useful for sequence and overlap determination are shown.

analysis. The MLA contains a glycosylation site at position 112 of the amino acid sequence. The sequence Asn^{112}-Gly^{113}-Ser^{114} could be identified as glycosylation pattern. The bound carbohydrate moiety showed a molecular weight of 1173.0 Da, resulting from the mass difference between the observed mass for the fraction T11 (MH^+ 2754.7 Da) and the calculated mass from the peptide sequence (MH^+ 1581.7 Da).

The alignment of the tryptic peptides was achieved by comparison with other homologous toxic lectins such as ricin and abrin [17,18] and confirmed by analysis of a set of overlapping peptides produced by cleavage with endoproteinase Asp-N. All peptides were completely sequenced except D11, where the first 39 amino acid residues could be identified. Fragments D11 and D14 precipitated after titration with HCl to pH 4.8, whereas all other peptides remained in solution at this pH value. Fig. 1 shows only the peptides which were useful in constructing the sequence.

However, while analysis of the peptides generated by endoproteinase Asp-N made it possible to confirm the predicted alignment of the tryptic peptides, otherwise obtained sequence analysis data from these two sets of peptides showed some

Table 1
Amino acid composition of intact MLA

Amino acid	Residues
Asp	22.2 (12)
Asn	(10)
Thr	20.1 (20)
Ser	23.3 (24)
Glu	28.2 (11)
Gln	(17)
Pro	13.6 (13)
Gly	16.1 (16)
Ala	17.0 (17)
½ Cys	n.d. (1)
Val	17.3 (17)
Met	2.7 (3)
Ile	17.3 (17)
Leu	25.0 (25)
Tyr	11.3 (11)
Phe	12.1 (13)
His	4.1 (4)
Lys	0.2 (0)
Arg	21.3 (21)
Trp	n.d. (2)
Total	(254)

Numbers in parentheses are the residues determined after sequence analysis. n.d., not determined.

heterogeneity. A comparison of the primary structure determined from the endoproteinase Asp-N digestion with that obtained from the tryptic cleavage reveals the presence of 17 conservative substitutions along the amino acid sequence. These two slightly different determined amino acid sequences can be due to the presence of two isoforms of the MLA. Observed replacements are summarized in Table 2. Some substitutions involve two structurally closely related amino acid sequence residues at the same position, such as replacements of $Glu^{15} \rightarrow Asp^{15}$, $Phe^{144} \rightarrow Tyr^{144}$ or $Thr^{231} \rightarrow Ser^{231}$, whereas in other cases no homology existed between them ($Thr^{151} \rightarrow Ala^{151}$ or $Ala^{184} \rightarrow Glu^{184}$). The most noteworthy substitution may be the replacement of $Asn^{112} \rightarrow Thr^{112}$, which supposes the elimination of the glycosylation signal in the amino acid sequence (Asn^{112}-Gly^{113}-$Ser^{114} \rightarrow Thr^{112}$-$Gly^{113}$-$Ser^{114}$). These results provide evidence that the A chain of ML-I occurs in at least two isoforms, one glycosylated (MLA) and one non-glycosylated (MLA').

4. Discussion

The goal of this work was to determine the primary structure of the cytotoxic chain of ML-I. Sequence information was obtained by analysis of two sets of peptides generated by trypsin and endoproteinase Asp-N, respectively. Collected fractions after RP-HPLC separation were examined by sequence and MALDI-MS analysis. The determined molecular masses of the fragments were in close agreement with those calculated from sequence analysis data (data not shown). MLA contains 254 amino acid residues. Examination of its primary structure shows the existence of only one half cystine residue (Cys^{247}) which presents evidence that MLA contains no internal disulfide bonds and the two chains of ML-I are linked by a single disulfide bond. Furthermore, the determined amino acid sequence of MLA encloses a potential site for asparagine-linked carbohydrates (Asn^{112}-Gly^{113}-Ser^{114}). The presence of covalently bound sugars at this residue was confirmed by mass spectrometric analysis. A molecular weight of 1173.0 Da could be determined for the carbohydrate chain.

The data presented here provide evidence that MLA also occurs in another isoform, non-glycosylated (MLA'). This finding correlates with a previous report [19], which suggested the existence of different isolectins of *Viscum album*. A comparison of the obtained primary structures indicates that these two isoforms differ at least at 17 positions. These sequence heterogeneities might not alter the properties of this protein significantly, since the amino acid residues in both proteins mostly have similar characteristics. Otherwise, carbohydrates in glycoproteins generally play no role in their catalytic function and therefore the absent carbohydrate chain in MLA' might not modify the glycosidic activity of the protein. In fact, it has been proved that the recombinant ricin A chain expressed in *E. coli* has full catalytic activity [20]. One possible role of carbohydrates in glycoproteins could be to increase their solubility in water. The presented amino acid sequences in this paper are very similar to those recently reported for the amino-terminal fragments of two toxic chains of ML-I isolated by two-dimensional SDS-PAGE [21].

The reason why the amino acid sequences of different isoforms (MLA and MLA') were determined is not clear, but it has been observed that lectin amounts in mistletoe preparations depend on the month of the mistletoe harvest [22]. Mistletoe extract composition can change according to the origin of the plant and the thickness of its leaves as well [23].

Finally, the amino acid sequence of the catalytic chain of ML-I was compared to those of other structurally and functionally closely related proteins like ricin and abrin [17,18]. These proteins are known as ribosome-inactivating proteins (type II RIPs) and share the same mechanism of action [5]. Although these plants are of a different phylogenetic origin, their toxic components present extensive structural similarities and operate as potent inhibitors of eukaryotic protein synthesis at the ribosomal level. The overall sequence comparison of the primary structure of the A chain of ML-I (MLA) with those of abrin-a (ABA) and ricin-D (RTA) shows a high degree of homology with 111 and 103 amino acid residues conserved, respectively (Fig. 2). N-terminal amino acid sequences are quite variable in comparison to some internal

Table 2
Substitutions among the amino acid sequences of the two isoforms MLA → MLA'

Glu^{15}	→ Asp^{15}
Val^{66}	→ Ile^{66}
*Asn^{112}	→ Thr^{112}
Pro^{114}	→ Thr^{114}
Asp^{133}	→ Glu^{133}
Gln^{134}	→ Glu^{134}
Thr^{140}	→ Ser^{140}
Phe^{144}	→ Tyr^{144}
Thr^{151}	→ Ala^{151}
Tyr^{179}	→ Asp^{181}
Ala^{184}	→ Glu^{184}
Val^{190}	→ Met^{190}
Ile^{218}	→ Phe^{218}
Pro^{223}	→ Ser^{223}
Pro^{224}	→ Thr^{224}
Thr^{231}	→ Ser^{231}
Asp^{235}	→ Ser^{235}

*This residue is the glycosylation site of MLA.

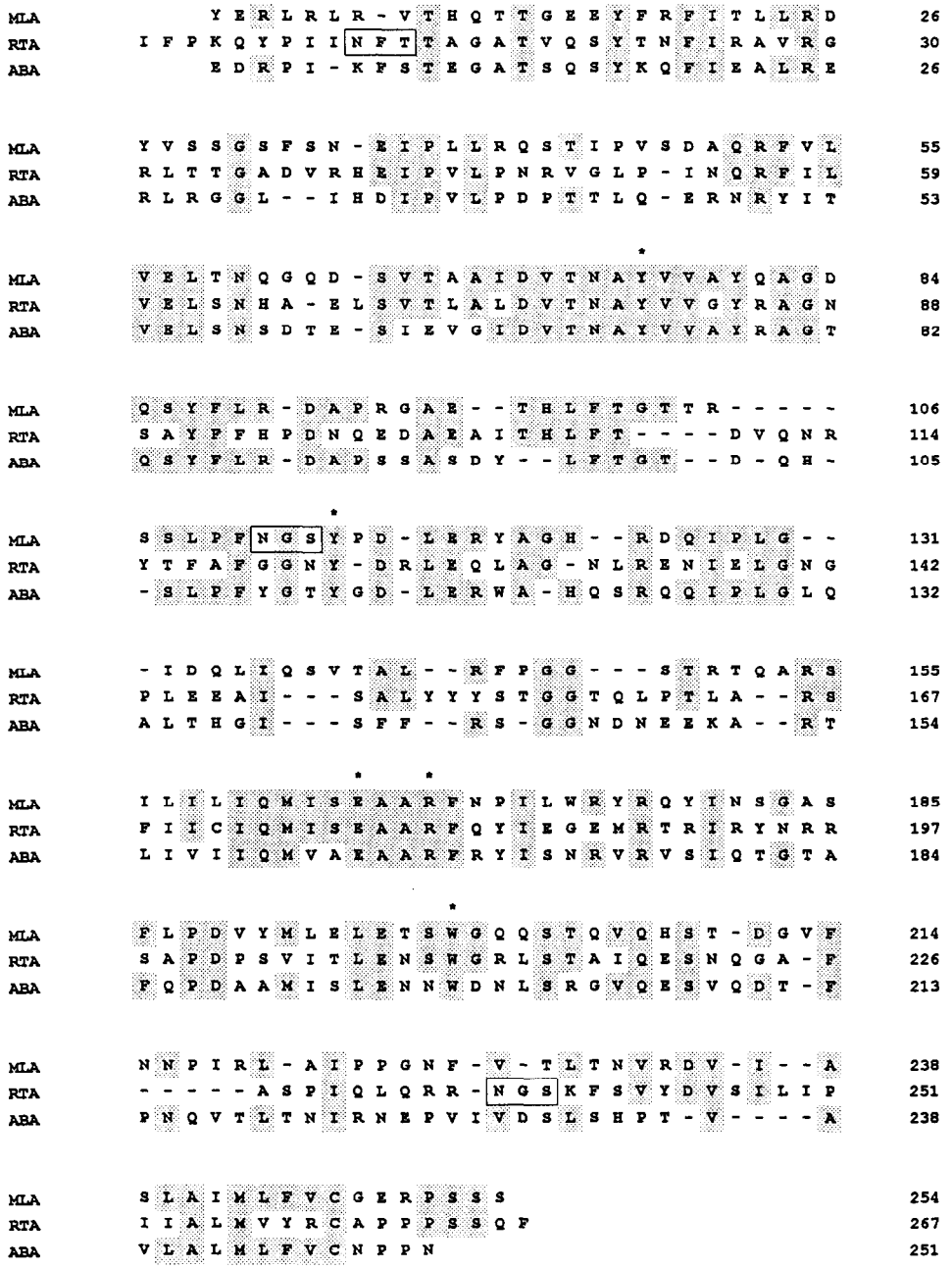

Fig. 2. Comparison of amino acid sequences of some ribosome-inactivating proteins (type II RIPs). MLA is the ML-I A chain. RTA is the ricin-D A chain [16] and ABA is the abrin-a A chain [17]. Sequences are aligned to maximize their similarities. Conserved residues among these proteins are highlighted. *N*-Glycosylation sites are boxed. Residues involved in the active site of the protein are marked with an * at the top of the column.

regions appearing as the most rigorously conserved parts of these molecules (over 80% identity between residues 69–83, for example).

A more extensive comparison after performing sequence alignment reveals that the amino acid sequence of the MLA conserves a number of absolutely invariant residues in the inhibitor protein family [24], which are presumably required for the catalytic function of the protein. These are Tyr^{76}, Tyr^{115}, Glu^{165}, Arg^{168} and Trp^{199} (Fig. 2). Other highly conserved residues, Asn^{74}, Arg^{125}, Gln^{161} and Glu^{196}, are also present. Their function remains to be elucidated, but they are believed to be less involved in maintaining the active site structure [25]. All residues mentioned above can also be identified in the amino acid sequence of the non-glycosylated form (MLA′). This seems to suggest that these functionally related proteins have arisen from a process of divergent evolution from a common ancestor, since they present an essentially identical arrangement of amino acid residues at the active site.

References

[1] Franz, H. (1985) in: Lectins: Biology, Biochemistry, Clinical Biochemistry (Bøg-Hansen, T.C. and Breborowicz, J., Eds.) Vol. IV, pp. 463–472, Walter de Gruyter, Berlin.
[2] Franz, H., Ziska, P. and Kindt, A. (1981) Biochem. J. 195, 481–484.
[3] Franz, H., Ziska, P. and Kindt, A. (1982) in: Lectins: Biology,

Biochemistry, Clinical Biochemistry (Bøg-Hansen, T.C., Ed.) Vol. II, pp. 171–176, Walter de Gruyter, Berlin.
[4] Endo, Y., Tsurugi, K. and Franz, H. (1988) FEBS Lett. 231, 378–380.
[5] Barbieri, L., Batelli, M.G. and Stirpe, F. (1993) Biochim. Biophys. Acta 1154, 237–282.
[6] Joule, R.J. and Neville, D.M. (1982) J. Biol. Chem. 257, 1598–1601.
[7] Beuth, J., Ko, H.L., Gabius, H.-J., Burrichter, H., Oette, K. and Pulverer, G. (1992) Clin. Invest. 70, 658–661.
[8] Hajto, T., Hostanska, K. and Gabius, H.-J. (1989) Cancer Res. 49, 4803–4808.
[9] Beuth, J., Ko, H.L., Tunggal, L., Geisel, J. and Pulverer, G. (1993) Arzneim.-Forsch./Drug Res. 43, 166–169.
[10] Hajto, T., Hostanska, K., Frei, K., Rordorf, C. and Gabius, H.-J. (1990) Cancer Res. 50, 3322–3326.
[11] Beuth, J., Ko, H.L., Tunggal, L., Buss, G., Jeljaszewicz, J. and Pulverer, G. (1994) Arzneim.-Forsch./Drug Res. 44 (II), Heft 11.
[12] Heiny, B.-M. and Beuth, J. (1994) Anticancer Res. 14, 1339–1342.
[13] Falconer, J., Chan, E.C. and Madsens, G. (1988) J. Endocrinol. 118, 5–8.
[14] Beuth, J., Ko, H.L. and Pulverer, G. (1994) Dtsch. Apoth. Z. 25, 17–28.
[15] Ziska, P., Franz, H. and Kindt, A. (1978) Experientia 34, 123–124.
[16] Olsnes, S., Stirpe, F., Sandvig, K. and Pihl, A. (1982) J. Biol. Chem. 257, 13263–13270.
[17] Halling, K.C., Halling, A.C., Murray, E.E., Ladin, B.F., Houston, L.L. and Weaver, R.F. (1985) Nucleic Acids Res. 13, 8019–8033.
[18] Hung, C.-H., Lee, M.-C., Lee, T.-C. and Lin, J.-Y. (1993) J. Mol. Biol. 229, 263–267.
[19] Juther, P., Theise, H., Chatterjee, B., Karduck, D. and Uhlenbruck, G. (1980) Int. J. Biochem. 11, 429–435.
[20] O'Hare, M., Roberts, L.M., Thorpe, P.E., Watson, G.J., Prior, B. and Lord, J.M. (1987) FEBS Lett. 216, 73–78.
[21] Gabius, S., Kaiser, K. and Gabius, H.-J. (1991) Dtsch. Zschr. Onkol. 5, 113–119.
[22] Franz, H., Ziska, P. and Kindt, A. (1985) in: Lectins: Biology, Biochemistry, Clinical Biochemistry (Bøg-Hansen, T.C. and Breborowicz, J. Eds.) Vol. IV, pp. 473–480, Walter de Gruyter, Berlin.
[23] Luther, P. and Becher, H. (1987) Die Mistel: Botanik, Lektine, medizinische Anwendung, Springer-Verlag, Berlin.
[24] Lord, J.M., Roberts, L.M. and Robertus, J.D. (1994) FASEB 8, 201–208.
[25] Betsy, J.K., Collins, E.J. and Robertus, J.D. (1991) Proteins Struct. Funct. Genet. 10, 251–259.