

## 2.8-Å Crystal Structure of a Nontoxic Type-II Ribosome-Inactivating Protein, Ebulin I

John M. Pascal,<sup>1</sup> Philip J. Day,<sup>1</sup> Arthur F. Monzingo,<sup>1</sup> Stephen R. Ernst,<sup>1</sup> Jon D. Robertus,<sup>1\*</sup> Rosario Iglesias,<sup>2</sup> Yolanda Pérez,<sup>2</sup> Jose Miguel Ferreras,<sup>2</sup> Lucia Citores,<sup>2</sup> and Tomás Girbés<sup>2</sup>

<sup>1</sup>Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas

<sup>2</sup>Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad de Valladolid, Valladolid, Spain

**ABSTRACT** Ebulin I is a type-II ribosome-inactivating protein (RIP) isolated from the leaves of *Sambucus ebulus* L. As with other type-II RIP, ebulin is a disulfide-linked heterodimer composed of a toxic A chain and a galactoside-specific lectin B chain. A normal level of ribosome-inactivating *N*-glycosidase activity, characteristic of the A chain of type-II RIP, has been demonstrated for ebulin I. However, ebulin is considered a nontoxic type-II RIP due to a reduced cytotoxicity on whole cells and animals as compared with other toxic type-II RIP like ricin. The molecular cloning, amino acid sequence, and the crystal structure of ebulin I are presented and compared with ricin. Ebulin I is shown to bind an A-chain substrate analogue, pteric acid, in the same manner as ricin. The galactoside-binding ability of ebulin I is demonstrated crystallographically with a complex of the B chain with galactose and with lactose. The negligible cytotoxicity of ebulin I is apparently due to a reduced affinity for galactosides. An altered mode of galactoside binding in the 2 $\gamma$  subdomain of the lectin B chain primarily causes the reduced affinity. *Proteins* 2001; 43:319–326. © 2001 Wiley-Liss, Inc.

**Key words:** plant toxin; *N*-glycosidase; galactoside lectin; cytotoxicity; ricin

### INTRODUCTION

Ribosome-inactivating proteins (RIP) are a well-characterized and ever-expanding class of proteins.<sup>1–4</sup> RIP from higher plants are divided into two classes. Type I RIP are single-chain toxins with *N*-glycosidase activity; they depurinate a specific adenine of ribosomal RNA essential for binding elongation factors.<sup>5</sup> Type-II RIP consist of an A chain with *N*-glycosidase activity and a galactoside-binding lectin B chain. The two chains are disulfide-linked to form a heterodimer with the ability to bind and enter cells. Upon delivery to the cytosol, the toxic moiety inactivates ribosomes, killing the cell. Whereas the A chain is responsible for the ribosome-inactivating activity, the B chain is responsible for attaching the toxin to cell surface galactosides and facilitating delivery to the cytosol. Both chains are required for maximum cytotoxicity.

Ricin, a 65-kD type-II RIP from *Ricinus communis*, is the archetype of the protein family. The structure, biochemistry, and cytotoxicity of ricin have been thoroughly examined and reviewed.<sup>6,7</sup> The key residues of the ricin A chain

(RTA) binding cleft have been identified, and a mechanism for the *N*-glycosidase activity has been described. Several substrate analogues have been bound to the active site cleft and their interactions viewed by crystallographic methods.<sup>8,9</sup>

The structure of ricin B chain (RTB) clearly revealed the mode of galactoside binding inherent in RIP B chains. The B chain has two distinct structural domains. Each of these domains contains three homologous subdomains designated  $\alpha$ ,  $\beta$ , and  $\gamma$ .<sup>10</sup> Only two of these subdomains, 1 $\alpha$  and 2 $\gamma$ , have shown galactoside-binding by crystallographic methods with ricin. Based on these two sites, three structural criteria for galactoside-binding have been proposed: an aromatic residue for stacking interaction with the nonpolar face of the galactose ring, a carboxylate for hydrogen bonding with galactose hydroxyls at positions 3 and 4, and a three residue kink in the polypeptide backbone that interacts with the polar face of the galactose ring.<sup>11</sup> The remaining four subdomains of RTB (1 $\beta$ , 1 $\gamma$ , 2 $\alpha$ , and 2 $\beta$ ) showed no bound galactosides, and each lacked at least one of the three structural elements described above.

Ebulin I is a 56-kD type-II RIP isolated from the leaves of *Sambucus ebulus*, commonly known as the dwarf elder.<sup>12</sup> Previous experiments have shown ebulin I A chain inactivates naked ribosomes with the same efficiency displayed by ricin.<sup>12,13</sup> Ebulin I is known to agglutinate red blood cells and is purified by binding to a galactoside affinity column.<sup>12</sup> However, ebulin I is ~300-fold less toxic to whole cells than is ricin, and is thus considered a nontoxic RIP. This raises the question of why ebulin I has such reduced cytotoxicity.

The molecular cloning, amino acid sequence, and crystal structure of ebulin I are presented here, and compared with ricin. Ebulin I A chain is shown binding to pteric acid (PTA) in the same manner as ricin, confirming that ebulin I has a functional A chain. Crystal complexes of ebulin I with galactose and with lactose demonstrate the mode of galactoside binding for ebulin in sites 1 $\alpha$  and 2 $\gamma$ . Ebulin I is shown to bind to an  $\alpha$ -D-lactose-agarose matrix with less

Grant sponsor: National Institutes of Health; Grant number: GM 30048; Grant sponsor: Foundation for Research and the Welch Foundation; Grant sponsor: CICYT; Grant number: BIO1998-0727.

\*Correspondence to: Jon D. Robertus, University of Texas at Austin, Welch Hall 5.266, Mail Code A5300, Austin, TX 78712. E-mail: jrobertus@mail.utexas.edu

Received 11 September 2000; Accepted 12 January 2001

affinity than ricin. The reduced cytotoxicity of ebulin is due to an impaired affinity for galactosides, caused by slight alterations in the  $\gamma$  subdomain of the B chain of the toxin.

## MATERIALS AND METHODS

### DNA Sequencing

DNA sequencing was carried out on the Aby Prism 377 automatic sequencer (PerkinElmer) at the Servicio de Secuenciación Automática de DNA (Centro de Investigaciones Biológicas, Madrid, Spain). The sequence reactions were performed with the Big-dye kit (PerkinElmer) using the pCR2.1 plasmid and the M13 and M13R universal primers, according to the manufacturer's recommendations.

### Synthesis of cDNA

Dwarf elder (*Sambucus ebulus* L.) young leaves from early summer were collected and immediately dropped into liquid nitrogen and stored until use. Total RNA was isolated by the guanidinium isothiocyanate method,<sup>14</sup> and poly(A)-rich RNA was purified by oligo-d(T) column chromatography. Poly(A)-rich RNA (0.2  $\mu$ g) was reverse transcribed using the synthetic oligonucleotide J1 (5'CGTCTAGAGTCTAGTAGTGC(T)<sub>20</sub>3') in a reaction mixture containing in 20  $\mu$ l: 1 mM of each dNTP, 5 mM MgCl<sub>2</sub>, 50 mM KCl, 20 mM Tris-HCl (pH 8.4), and 20 U of pMuLV reverse transcriptase (PerkinElmer). The reaction mixture was incubated for 16 min at 23°C (room temperature), then for 20 min at 42°C, and finally 5 min at 99°C.

### PCR Amplification of cDNA

For polymerase chain reaction (PCR) amplification, two oligonucleotide primers were synthesized—sense A1 (5'AATTTGGCGGGTGCCAAARWSIAC3'), and antisense J2 (5'CGTCTAGAGTCTAGTAGTGC3'). A1 was synthesized based on the N-terminal amino acid sequence of ebulin 1.<sup>12</sup> In this investigation, 2  $\mu$ l of the above-synthesized cDNA was amplified by PCR (95°C 30 s, 50°C 1 min 30 s, 72°C 2 min, for 40 cycles) in the presence of primers A1 and J2. The amplified DNA product was purified by the High Pure PCR Product Purification kit (Boehringer-Mannheim), cloned into the pCR2.1 vector (Original TA cloning kit, Invitrogen) and sequenced in both directions.

Full-length cDNA was obtained by the 5'RACE system from Gibco-BRL (5'RACE system for rapid amplification of cDNA ends, version 2.0). First-strand cDNA was synthesized from 0.2  $\mu$ g poly(A<sup>+</sup>) RNA using a gene-specific primer B3 (5'GATGACGATCAGATCCTTGAGAC3'). After first-strand cDNA synthesis, the original mRNA template is removed by treatment with a RNase Mix (mixture of RNase H and RNase T1 from Gibco-BRL). Unincorporated dNTP, primer, and proteins were separated from cDNA using a Glassmax spin cartridge (Gibco-BRL). A homopolymeric tail was then added to the 3' end of the cDNA using terminal transferase enzyme (TdT) and dCTP. PCR amplification is accomplished using 2.5 U of Amplitaq DNA polymerase (Perkin-Elmer, Branchburg, NJ), a nested gene-specific primer B2 (5'GTTCGATAGACTCCTCC3') that anneals to a site located within the cDNA

molecule, and a deoxyinosine-containing anchor primer AAP (5'GGCCACGCGTCTAGTAGTACGGGIIGGGIIGGGIIG3') provided with the AAP System (Abridged Anchor Primer, Gibco-BRL).

The primary PCR product was reamplified using a gene-specific primer B1 (5'AGATAACGTATGTTGCGTGTGTC3') and the abridge universal primer AUAP (5'GGC-CACGCGTCTAGTAGTAC3') provided with the AUAP System (Abridged Anchor Primer, Gibco-BRL). 5'RACE product was cloned into pCR2.1 vector (Invitrogen) and then sequenced.

### cDNA Cloning

The cDNA from ebulin 1 mRNA was cloned using the 3'RACE technique with a strategy similar to that used by Wu et al.<sup>15</sup> A degenerated primer specific for the N-terminal stretch of ebulin A chain (primer A1) and an unspecific primer for binding to the poly(A) tail (primer J1) were constructed. This permitted amplification of a cDNA fragment that contained most of the ebulin sequence (A chain, B chain, and 3'UTR). The leader and the 5'UTR sequences were determined by 5'RACE technique as follows.<sup>16</sup> A first copy of cDNA was obtained by RT using a specific primer from a well-known internal sequence of the B chain. Next a poly(C) tail was added to the 5' of this cDNA, which enabled the binding of commercial primers AAP and AUAP. Then, two amplifications were carried out: one with the sense primer AAP and the antisense primer B2 and the other one with the sense AUAP and the antisense B1. Two fragments of cDNA were obtained containing both the leader and the 5'UTR sequences. Such fragments overlapped with the cDNA sequence obtained by mean 3'RACE technique.

### Crystallization

Ebulin 1 was purified as previously reported.<sup>12</sup> Lyophilized protein was suspended in 10 mM Tris-HCl (pH 7.5) at concentrations ranging from 3 to 15 mg/ml. Hampton's Crystal Screen I (Hampton Research, Laguna Niguel, CA) yielded suitable leads for diffraction-quality crystals. The final crystallization conditions of the two crystal forms used in structure determination are as follows: (1) 1.0 M Li<sub>2</sub>SO<sub>4</sub>, 2% PEG 8000, 100 mM Tris-HCl pH 8.5:10–15 mg/ml protein, 4°C (orthorhombic); and (2) 1.1 M Na,K tartrate, 100 mM Hepes pH 7.5:3 mg/ml protein, 25°C (trigonal). Both orthorhombic and trigonal crystal forms were grown by hanging drop vapor diffusion (1:1, precipitating solution to protein). Trigonal crystals were exposed to the following compounds by adding each compound to a volume of artificial mother liquor in which the crystal was soaked in (1) 2 mM lactose for 3 days; (2) 2 mM galactose for 3 days; and (3) a saturated solution of pteric acid for 5 days.

### Data Collection and Reduction

Crystals were exposed to X-rays from a Rigaku RU200 rotating anode generator operated at 50 kV and 100 mA. Data were collected on a Rigaku Raxis IV image plate detector (Molecular Structure Corporation, Woodlands,

TX). Orthorhombic crystals were capillary-mounted and X-rayed at room temperature. Trigonal crystals were transferred briefly (<1 s) to a 1:1 solution of artificial mother liquor to 60% glycerol and then immediately into liquid nitrogen. The cryo-cooled trigonal crystals were maintained at  $-120^{\circ}\text{C}$  during data collection. The HKL suite<sup>17</sup> was used to index and reduce the data. The two crystal forms were determined to belong to the orthorhombic space group  $P2_12_12_1$  and the trigonal space group  $P3_221$  or  $P3_121$ , respectively. Data from three orthorhombic crystals were merged into a single native data set.

### Molecular Replacement, Model Building, and Refinement

Molecular replacement was carried out with the orthorhombic native data and with ricin as a model<sup>18</sup> (PDB 2AAI) using X-PLOR.<sup>19</sup> The highest peak from the rotation search was  $4.0\sigma$  above the mean; the next highest being  $3.3\sigma$  above the mean. The Patterson correlation coefficient of the highest peak was 2.5 times larger than the next highest peak. The highest peak from the translation search was  $8.3\sigma$  above the mean, and the next highest peak was only  $4.8\sigma$  above the mean. At this point, all residues in the ricin model were mutated to alanine, except conserved residues that were strongly represented in electron-density maps. The ebulin I sequence was added through successive rounds of model refinement and rebuilding. Loops with residues inserted or deleted were adjusted as polyalanines until side-chain electron density was apparent. Two types of weighting schemes were used to alleviate model bias in  $2\text{Fo}-\text{Fc}$  electron-density maps: SIGMAA from the CCP4 suite<sup>20</sup> and LBEST.<sup>21,22</sup> Model coordinates and density maps were visualized on a Silicon Graphics Indy using O.<sup>23</sup> X-PLOR<sup>24-26</sup> was used to perform simulated-annealing model refinement, bulk solvent data correction, and selection of ordered water molecules. Later rounds of refinement included cycles of PROTIN and REFMAC from CCP4.<sup>27</sup>

The refined orthorhombic model was used as the search model for a molecular replacement search of the trigonal ebulin-lactose data, using EPMR (Agouron Pharmaceuticals, La Jolla, CA). A clear solution (0.643 correlation coefficient, 0.392 R) was found when performing the search in space group  $P3_221$ , indicating that to be the correct enantiomorphic space group. X-PLOR maps<sup>28,29</sup> of the type  $\text{Fo}-\text{Fc}$  and  $2\text{Fo}-\text{Fc}$  with free R cross-validated SigmaA weighting were most helpful in determining structural changes in the trigonal crystal form and in locating bound lactose, galactose, and pteric acid. The geometric and stereochemical quality of the models was assessed using PROCHECK.<sup>30</sup>

### $\alpha$ -D-Lactose-Agarose Affinity Columns

Ricin and ebulin were tested for their relative binding affinity to  $\alpha$ -D-lactose conjugated to agarose beads (Sigma). A laboratory supply of ricin from plant seeds was used.<sup>18</sup> Both toxins were exhaustively dialyzed into phosphate-buffered saline (PBS), and then 1 ml of each protein ( $\sim 0.5$  mg/ml) was loaded onto separate 1-ml  $\alpha$ -D-lactose-agarose

columns. Each column was then washed with three-column volumes of PBS. D-Galactose was used to elute bound protein by competition with the galactose moiety of  $\alpha$ -D-lactose; 2-ml volumes of PBS containing increasing levels of D-galactose (2.5–900 mM) were applied to the column in a step gradient. The binding/washing eluate and each of the eluting steps were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The bands from SDS-PAGE were digitized and then integrated using ImageJ (Wayne Rasband, NIH).

## RESULTS AND DISCUSSION

The cDNA for ebulin I contains a 1,692-bp open reading frame (ORF) encoding a polypeptide of 564 amino acids, with one possible initiation codon at position  $-25$  of the processed expressed amino acid sequence. The gene sequence has been deposited in the EMBL Nucleotide Sequence Database, accession number AJ400822. Hydrolytic processing of the signal peptide between residues  $-1$  and  $1$  results in a polypeptide of 539 amino acids that contains the N-terminal amino acid sequence<sup>12</sup> of both the A and B chains of ebulin I. Additional processing removes the linker peptide and generates the two disulfide-linked ebulin chains. A comparison of the amino acid sequence of ebulin with ricin (Fig. 1) indicates the A chains share 34% amino acid identity and the B chains share 48% identity.

Table I summarizes crystallographic data collected for the two forms of ebulin. Table II summarizes the refinement data for both the orthorhombic and trigonal models of ebulin. Figure 2 exhibits representative electron density for a section of the 2.8-Å electron-density map for the A chain of the trigonal form.

The two refined models have very similar folding, but there are significant structural differences between the orthorhombic and trigonal forms. The orientation for residues A147–A149, B156–B160, and B239–B243 is different for the two crystal forms. In the orthorhombic model, B-chain residues 156–160 and 239–243 contact residues A147–A149 of a neighboring molecule. The root-mean-square deviation (RMSD) for all  $\text{C}_\alpha$  atoms is 0.96 Å, but excluding the above-mentioned lattice contact areas, where  $\text{C}_\alpha$  atoms differ in positioning by as much as 6 Å, the RMSD for the remaining  $\text{C}_\alpha$  atoms is 0.64 Å. Lattice contacts with these B-chain residues of the orthorhombic crystal form distort sugar-binding subdomain  $2\gamma$ . Therefore, the trigonal crystals, lacking this symmetry contact, are the better choice for examining sugar binding.

Figure 3 is a ribbon drawing of the backbone for the ebulin heterodimer solved from the trigonal crystal data. Ebulin has essentially the same secondary and tertiary structure as ricin. To illustrate this, Figure 4 shows a least-squares superposition of the trigonal ebulin model and ricin. Both A and B chains are clearly homologues. The RMSD of the  $\text{C}_\alpha$  for the A chains is 1.51 Å and 1.36 Å for the two B chains.

Although the ricin A chain is slightly larger than the ebulin A chain (267 residues versus 253), there are no major structural differences. The ebulin A model has two sizeable deletions, compared with ricin A, which accounts

## A chains

```

          10      20      30      40      50
ebulin -I--DYPSVSFNLGAKSTTYRDFLNLRDRVATGTVEVNGLPVLRRESEVQVKNRFVLV
      . . . . .
ricin  IFFKQYPIINFTTAGATVQSYTNFIRAVRGRLLTGADVRHEIPVLPNRVGLPINCQRFILV

          60      70      80      90     100     110
ebulin RLTYNNGQDTVTSVAIDVTNLVVAFAFSAANGNSYFEKDA---TELQKSNLPLGTFQHTLSE
      . . . . .
ricin  ELSNHAELSVTALDVTNAYVGVYRAGNSAYFFHPDNQEDAEAI--TLLPTDVQNRVYTFAF

          120     130     140     150     160
ebulin TGNVYDNIETAAAGTRRSTFELGPNPLDGAITSWYD)-----GGVARSLIWTIQMVFPEAA
      . . . . .
ricin  CGNYDRLEQLAGNLRNLELGNGLPEEAIAISALYYYSYGGTQLPILARSLIICIQMISEAA

          170     180     190     200     210     220
ebulin RFRYLEQEVRRSLQQLTSFTPNALMLSMENNWSMSLEVLQSGDNVSPFSGTVQLQNYDH
      . . . . .
ricin  RFQYIEGEMRTRTRYNRRSAPPDSVTTFRNSWGRISTATQFSN--QGAFASPIQIQRRNG

          230     240     250
ebulin TPRLVDNFEELYKIPTGIALLLFRCVATKT---
      . . . . .
ricin  SKFSVYDVSILIP--IIALMVYRCAPPSSSQF
          240     250     260

```

## B chains

```

          10      20      30      40      50
ebulin DGECALPAPFTRRIVGRDGLCVDVVRNGYDTDGTPIQLWPCG--TQRNQGWVYNDKTR
      . . . . .
ricin  -ADVCMDP--EPIVRIVGRNGLCVDVVRDGRFENGNAIQWPCSKSNTDANQLWVNLKRDNTIR

          60      70      80      90     100     110
ebulin SMGKCMTANGLNSGSYIMITDCSTAADATKWEVLIDGSIINPSSGLVMTAPSGASRTTL
      . . . . .
ricin  SNGKCLITYGSPGVYMIYDNCNTAATDTRWQIWDNGTIIINPRSSLVLAATSQNSGTTL

          120     130     140     150     160     170
ebulin LLENNIHAASQGWTVSNDVQPIATLGVYNEMLQANGENNNVWMECDVTVSVQQQWALF
      . . . . .
ricin  TVQNTIYAVSQGWLPNTNTPQFVTTIVGLYGLCLQAN--SGQVWIEDCSSEKAEQQWALY

          180     190     200     210     220     230
ebulin DDRTRVNNRSGLCVTSNGYVSKDLVIRKCCQLLATQRWFNSDCSVNLRKSTRVMDVK
      . . . . .
ricin  ADGSRPQQNRNCLTSDSNIRETVVKIILSCGPASSGQRWMEFKNDGTTILNLYSGLVLDVLR

          240     250     260
ebulin ESDVSLQEVIIIFPAGNPNQWRTQVPPQI
      . . . . .
ricin  ASDPSLKQIILLYPLHGDPNQLWLPF---
          240     250     260

```

Fig. 1. Amino acid sequence alignment of A and B chains of ricin and ebulin. Sequence identities are marked with double dots (:). Sequence conservations are marked with single dots (.). Key residues of the A-chain active site and the B-chain galactoside-binding clefts are in bold.

for most of the size discrepancy. One deletion is a surface loop that is 7 residues shorter in ebulin A, equivalent to the loss of ricin A chain residues 155–161. The other is also a surface loop that is shortened four residues in ebulin, corresponding to ricin residues 97–100. The positions of these two loops are indicated in Figure 4. Two other areas of ebulin A chain have notable deviations in the polypeptide chain compared with RTA. Both arise from ebulin inserting two residues. Ebulin residues 209 and 210 are inserted in a surface loop. Residues 239 and 240 extend an  $\alpha$ -helical region of ricin (residues 245–248) by two residues.

Ebulin has roughly the same positioning of key active site residues as the ricin A chain. One exception is that the side-chain of the Tyr 77 (Tyr 80 in RTA) of ebulin is rotated out of the binding site pocket, in an orientation more similar to that seen in PAP, a type I RIP.<sup>31</sup> The mode of substrate binding for the active site of ricin has been established by analyzing complexes with the substrate

analogue FMP<sup>8</sup> and the inhibitor PTA.<sup>9</sup> We soaked PTA into the trigonal ebulin crystals and found it binds in the same manner as for ricin (Fig. 5). The pterin ring stacks with the side-chain of Tyr 77 and forms hydrogen bonds with the backbone of Leu 78 and Gly 114. Arg 166 donates a hydrogen bond to N<sup>5</sup> of the pterin. This active site structural similarity is consistent with the fact that ebulin displays essentially the same enzyme activity as ricin in a cell-free protein synthesis assay.<sup>12</sup> Since the activity of ebulin and ricin A chains is so similar, differential toxicity between the heterodimeric toxins, observed on intact animal cells, is most likely due to differences in their B chains.<sup>13,32</sup>

The overall fold of the ebulin B chain is very similar to that of the ricin B chain; there are only minimal deviations from RTB in the trace of the polypeptide chain of ebulin (Fig. 4). Ebulin has two areas in which residues are inserted, as compared with ricin: a one-residue insertion near the N-terminus (Ala 9) and a two-residue insertion in subdomain 1 $\gamma$  (Gly 156 and Glu 157). There are two deletions of ebulin B chain compared with RTB: a two-residue deletion in subdomain 1 $\alpha$  (ricin residues 41 and 42), and a deletion of ricin residue Gly 208. None of the insertions or deletions has a dramatic effect on the fold of the protein. The insertion within subdomain 1 $\alpha$  (sugar-binding) does not affect the geometry of residues known to be involved in lectin function. There are no deletions or insertions within subdomain 2 $\gamma$ , the other known sugar-binding domain.

Ricin crystals bind the disaccharide lactose in two of the six B-chain subdomains, 1 $\alpha$  and 2 $\gamma$ . As shown in Figure 6, trigonal ebulin binds lactose in its 1 $\alpha$  subdomain in a nearly identical fashion to ricin. The ebulin 1 $\alpha$  subdomain uses the same sugar-binding residues as RTB; Trp 39, Asp 24, Gln 37, Asn 46, and Gln 47 of ebulin all interact with the galactose unit. These ebulin residues correspond to ricin residues Trp 37, Asp 22, Gln 35, Gln 46, and Asn 47. No lactose was bound in the ebulin 2 $\gamma$  subdomain, although ricin and ebulin have very similar 2 $\gamma$  binding site geometries; that is, ricin residues Tyr 248, Asp 234, Asn 255, and Gln 256 correspond to Phe 249, Asp 235, Asn 256, and Gln 257 in ebulin.

When a galactose complex was formed, ebulin again showed the same mode of sugar-binding in the 1 $\alpha$  subdomain as ricin (not shown). Galactose is also bound in the 2 $\gamma$  subdomain of ebulin. Although the side-chain alignment of ricin and ebulin seems quite similar, the mode of galactose binding in the subdomain 2 $\gamma$  of ebulin is somewhat different (Fig. 7). First, the orientation of the galactose is rotated in the plane of the molecule nearly 90° from that of the galactose moiety in ricin. This changes the hydrogen-bonding pattern between the galactose hydroxyls and the charged residues at the top of the binding cleft. In ricin, the hydroxyls of C3 and C4 are within strong hydrogen bonding distance of residues Asp 234 and Asn 255. Asp 234 seems the predominant hydrogen-bonding partner for the galactose moiety. For ebulin, the C2 and C3 hydroxyls are in hydrogen-bonding distance of Asp 235 and Asn 256, and Asn 256 appears to dominate the galactose interaction.

TABLE I. Crystal Data Statistics

	Orthorhombic	Trigonal Lactose	Trigonal Galactose	Trigonal PTA
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P3 <sub>2</sub> 21	P3 <sub>2</sub> 21	P3 <sub>2</sub> 21
Cell parameters (Å)	<i>a</i> = 57.3 <i>b</i> = 97.7 <i>c</i> = 114.0	<i>a</i> = <i>b</i> = 74.0 <i>c</i> = 191.5	<i>a</i> = <i>b</i> = 73.6 <i>c</i> = 190.9	<i>a</i> = <i>b</i> = 74.2 <i>c</i> = 191.0
Resolution limit (Å)	2.8	2.9	2.8	3.1
Number of crystals	3	1	1	1
Completeness (%)	86.0	97.7	99.8	97.5
Unique reflections	13,477	12,522	15,454	11,311
<i>I</i> /σ <sub>1</sub> (last shell)	8.9 (5.3)	11.5 (3.0)	11.6 (2.3)	7.3 (2.1)
<i>R</i> <sub>merge</sub> (last shell) (%)	9.3 (32.1)	8.7 (28.6)	8.9 (38.3)	11.8 (37.8)

PTA, pteric acid.

TABLE II. Refined Model Statistics

	Orthorhombic	Trigonal Lactose	Trigonal Galactose	Trigonal PTA
Residues				
A chain	1–251	1–252	1–254	1–252
B chain	3–266	2–264	2–264	2–264
Galactose	0	0	2	0
Lactose	0	1	0	1
Glycosylation at B185 Asn	2 NAG, 3 MAN	2 NAG, 1 MAN	2 NAG	2 NAG, 1 MAN
Bound waters	56	37	46	20
Average temperature factors (Å <sup>2</sup> )				
All protein atoms	20.65	27.67	28.37	26.72
Waters	18.67	18.90	24.07	19.54
Side-chains	23.30	28.04	29.05	27.43
Resolution range (Å)	30.0–2.8	30.0–2.9	30.0–2.8	40.0–3.1
Crystallographic <i>R</i>				
All refs with <i>F</i> > 2σ <i>F</i>	0.194	0.237	0.264	0.219
Without bound waters	0.216	0.254	0.279	0.229
<i>R</i> <sub>free</sub>				
All refs with <i>F</i> > 2σ <i>F</i>	0.285	0.339	0.328	0.317
Without bound waters	0.291	0.344	0.336	0.325
RMSD				
Bond distances (Å)	0.004	0.003	0.003	0.005
Bond angles (°)	1.370	0.658	0.653	1.054
Dihedral angles (°)	30.771	27.685	28.304	30.982
Planarity (°)	0.465	0.423	0.462	0.400
Ramachandran plot				
Favored regions (%)	86.4	72.4	76.0	74.8
Disallowed residues	0	0	0	0

PTA, pteric acid; RMSD, root-mean-square deviation.

Asp 235 is oriented in a slightly different direction, away from the saccharide. Second, the positioning of galactose within the binding cleft is translationally shifted as compared with ricin. Galactose bound to ebulin is ~2.5 Å further into the binding cleft than the galactose moiety bound to ricin. Furthermore, the galactose in ebulin is shifted away from the hydrophobic stacking partner (Phe 249), toward the polypeptide kink, and slightly twisted out of the pocket toward Asn 254. In the polypeptide kink, ricin residue Ala 237 is changed to Glu 238 in ebulin. That residue bonds to the C6 hydroxyl of galactose in its altered orientation. Because of this orientation, galactose bound to ebulin has stronger

hydrogen-bonding interaction with the backbone of the polypeptide kink than that in ricin. The amine of Glu 238 forms a hydrogen bond with the O<sup>4</sup> hydroxyl of galactose, and the amine of Ser 239 also forms a hydrogen bond with the O<sup>4</sup> hydroxyl of galactose.

The orientation of galactose in subdomain 2γ of ebulin is subtly, but distinctly, different from that seen in ricin. The consequence of this orientation is that it directs the C1 hydroxyl toward Asn 254. This would cause steric interference for any sugars attached to the C1 hydroxyl, and is presumably why no lactose was found in site 2γ of ebulin crystals. This altered mode of galactose binding in the 2γ site of ebulin may indicate a weaker binding to complex

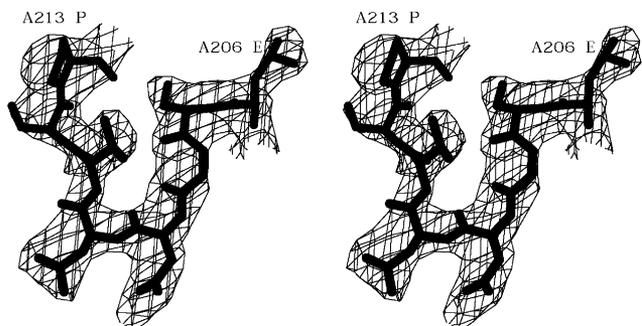


Fig. 2. Section of the 2.8-Å electron density for the trigonal model. A 2Fo-Fc map with sigmaA-weighted phases. The map is contoured at 1.3  $\sigma$ .



Fig. 3. Ribbon drawing of the galactose-bound model of ebullin in stereo. The termini of the A and B chains are labeled. Bound galactose residues are drawn in thick bonds. Drawing was made with MOLSCRIPT.<sup>40</sup>

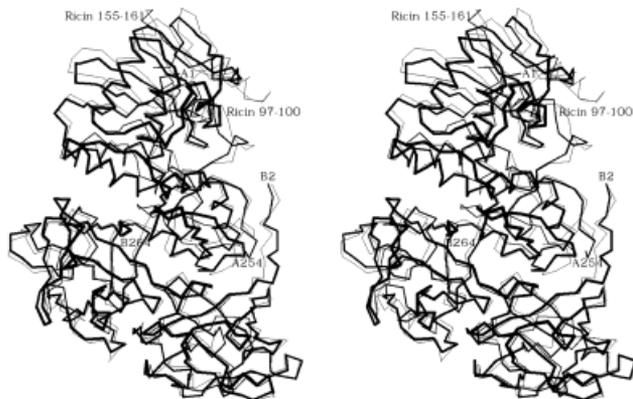


Fig. 4. Superposition of a C $\alpha$  trace of ebullin 1 and ricin in stereo. Ebullin is drawn in thick lines; ricin is drawn in thin lines. The A- and B-chain termini of ebullin are labeled. The ricin numbering for two deletions in the A chain are marked. Least-squares superposition of molecules was done with O.<sup>23</sup>

sugars. This, in turn, may alter or diminish binding to cell surfaces.

As there are no significant structural changes between ricin and ebullin within the 2 $\gamma$  subdomain, the change from Tyr to Phe at position 249 may account for the different binding mode seen in the two proteins. There are other nontoxic type-II RIP for which sequence, but not struc-

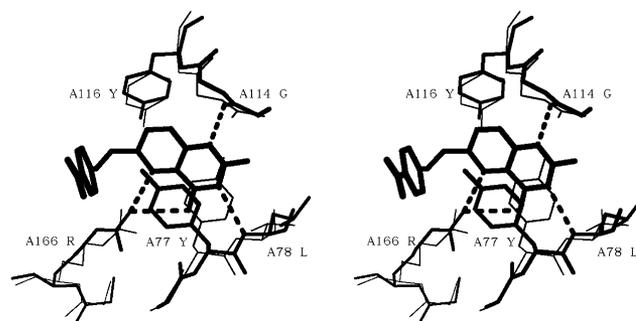


Fig. 5. Stereo view of pteric acid in the ebullin active site with key residues labeled. Analogous residues of ricin have been superimposed in thin lines.

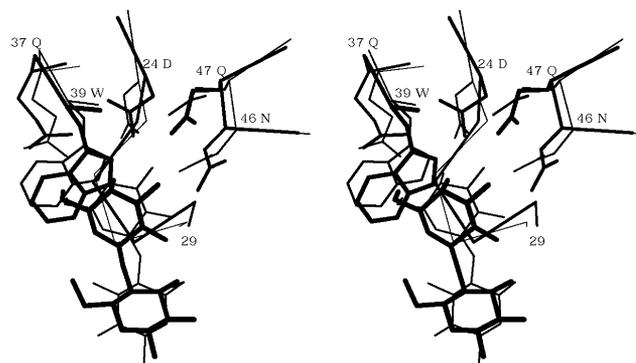


Fig. 6. Galactoside-binding subdomain 1 $\alpha$  of ebullin compared with that of ricin. The key residues in ebullin are shown in thick lines with the corresponding residues of ricin superimposed in thin lines. The positioning of a lactose molecule is shown for both ebullin (thick lines) and ricin (thin lines).

tural, information is available. Viewing the aligned sequences of nigrin b and sieboldin b shows a Phe at their positions relative to Phe 249 of ebullin.<sup>33,34</sup> The hydrogen bonding between the O<sup>6</sup> hydroxyl of galactose and Glu 238 in ebullin may be adventitious, a result of other changes within site 2 $\gamma$ . Nigrin b also has Glu at this position, but sieboldin b has Ala, as does ricin. It is not readily apparent from the structure of ebullin why the mode of galactose binding has changed for subdomain 2 $\gamma$ . Slight changes in the arrangement of key residues within the 2 $\gamma$  subdomain could cause the galactose-binding mode to shift to another preferred orientation. Why a Tyr to Phe substitution would alter binding of galactosides is also unclear. However, conservation of the substitution among other nontoxic, type-II RIP makes a strong case for the Phe having a significant role in the altered binding mode.

To compare the affinities of ebullin and ricin for a dense matrix of sugars, both ricin and ebullin were separately passed over lactose-agarose affinity columns. As shown in Figure 8, ricin eluted at a significantly higher concentration of free galactose. This result supports the idea that ebullin has some defect that limits its ability to bind polysaccharides terminating in galactose. Of the two binding sites in ricin, the 2 $\gamma$  galactose-binding site is the higher affinity site.<sup>35</sup> Perturbation of this site in ebullin could

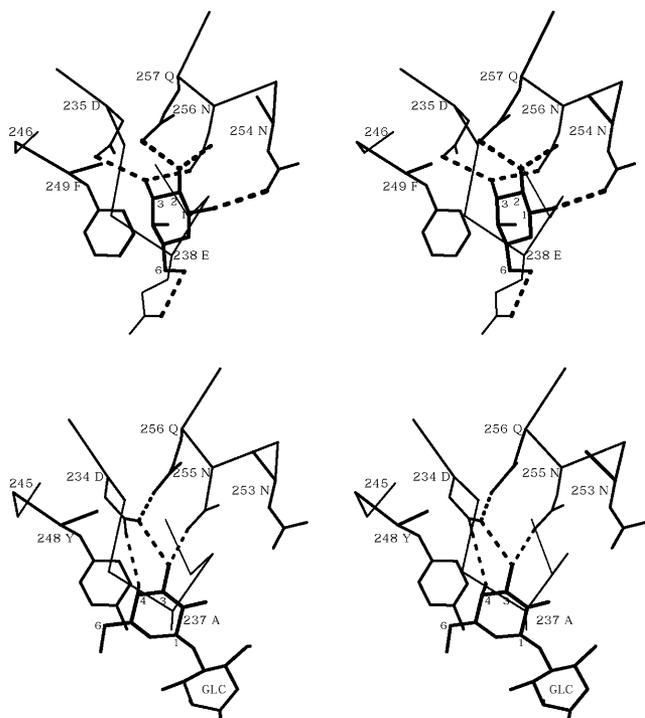


Fig. 7. Comparison of the galactoside-binding,  $2\gamma$  subdomains of ebulin (top) and ricin (bottom). Galactose is bound to ebulin. Lactose is bound to ricin. In both pictures, the galactose moiety is numbered at significant carbons of the sugar ring. All contacts with lactose in ricin are made with the galactose moiety. The glucose moiety (GLC) extends into the solvent. The GLC extension ( $\beta$ -1,4 linkage) from the C1 hydroxyl of the galactose moiety in ricin is not possible for the C1 hydroxyl of galactose in ebulin, because of steric interference with 254 N.

therefore lead to the observed drastic reduction in galactose affinity.

Recent studies have raised the possibility of a third sugar-binding site on ricin B chain. Mutational studies have shown that ricin can still bind galactosides modestly even if the confirmed  $1\alpha$  and  $2\gamma$  sites are partially disrupted.<sup>36</sup> Affinity-directed chemical modifiers interact with ricin B chain at subdomain  $1\beta$  (in proximity of Tyr 78), as well as within the confirmed subdomains  $1\alpha$  and  $2\gamma$ .<sup>37</sup> Tyr 78 is analogous to the hydrophobic stacking residue of the confirmed sugar-binding domains. However, subdomain  $1\beta$  lacks the conserved, charged residues responsible for hydrogen bonding to galactose hydroxyls, and no saccharide binding has been observed for site  $1\beta$  in ricin (nor in ebulin). Despite this, it is conceivable that the Tyr at position 78 is accountable for some of the galactoside affinity of ricin. By contrast, ebulin has Thr 78 at this same position. The lack of a hydrophobic residue at this location should entirely abolish any potential saccharide interaction. It is possible that this may contribute to the reduced ebulin affinity for galactoside matrices. However, we consider the alterations in subdomain  $2\gamma$  to be largely responsible for the observed reduced affinity of ebulin.

*Ricinus communis* produces isoforms of ricin with a lower affinity for galactosides. The common isoform, discussed above, is known as ricin D. *Ricinus communis*

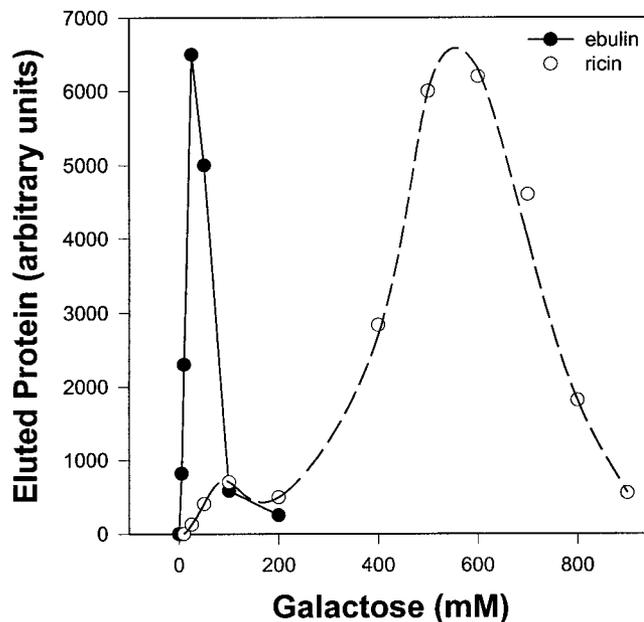


Fig. 8. Profiles of ebulin and ricin eluted with galactose from  $\alpha$ -lactose-agarose affinity columns.

agglutinin (RCA) is an isoform that associates as a heterotetramer of the form  $A_2B_2$ . However, due to a His substitution for Tyr 248 in the  $2\gamma$  subdomain, only the  $1\alpha$  subdomain of RCA binds galactosides with high affinity. RCA still has the ability to agglutinate cells because the module provides two  $1\alpha$  sugar-binding sites. Another isoform, ricin E, is apparently the result of a gene recombination of ricin D and RCA.<sup>38</sup> In the B chain, domain 1 is from ricin D and thus contains the functional  $1\alpha$  sugar-binding site. Domain 2 is from RCA and does not bind galactosides with high affinity.<sup>39</sup> Ebulin I might be a functional homologue of ricin E. What use *Sambucus ebulus* and other plants have for nontoxic, type-II RIP remains unclear. The carbohydrates and cell types we have examined may not represent the natural B chain ligands. It may be that the true binding target of ebulin, and proteins of its class, has not been discovered.

## CONCLUSIONS

We have shown that ebulin crystals bind galactose and lactose in subdomain  $1\alpha$  in a similar manner to that of ricin. Although ricin binds polysaccharides terminating in galactose in subdomain  $2\gamma$ , ebulin crystals bind only the monosaccharide galactose, and not the disaccharide lactose. Galactose bound to subdomain  $2\gamma$  of ebulin is oriented in a fashion that would prevent further saccharides from being attached to the C1 hydroxyl. This is presumably why ebulin crystals did not bind lactose in subdomain  $2\gamma$ . We have also shown that ebulin has a lower affinity than ricin for a dense matrix of lactose. The reduced cytotoxicity of ebulin compared with other type-II RIP is most likely due to a weaker affinity for terminal galactosides on cell surfaces. An altered mode of saccharide binding in subdomain  $2\gamma$  is responsible for the reduced affinity.

## ACKNOWLEDGMENTS

This work was supported by grant GM 30048 from the National Institutes of Health, by grants from the Foundation for Research and the Welch Foundation, and by grant BIO1998-0727 from Comisión Interministerial de Ciencia y Tecnología (to T.G.).

## REFERENCES

- Olsnes S, Pihl A. Toxic lectins and related proteins. In: Cohen P, Van Heynigen S, editors. The molecular action of toxins and viruses. New York: Elsevier; 1982. p 52–105.
- Jimenez A, Vazquez D. Plant and fungal protein and glycoprotein toxins inhibiting eukaryote protein synthesis. *Annu Rev Microbiol* 1985;39:649–672.
- Stirpe F, Barbieri L. Ribosome-inactivating proteins up to date. *FEBS Lett* 1985;195:18.
- Girbés T, Ferreras JM. Ribosome-inactivating proteins from plants. *Recent Res Dev Agric Biol Chem* 1998;2:1–16.
- Endo Y, Mitsui K, Motiguki M, Tsurugi K. The mechanism of ricin and related toxins on eukaryotic ribosomes. *J Biol Chem* 1987;262:5908–5912.
- Lord JM, Roberts LM, Robertus JD. Ricin: structure, mode of action, and some current applications. *FASEB J* 1994;8:201–208.
- Robertus JD, Monzingo AF. The structure of ribosome inactivating proteins from plants. In: Parker M, editor. Protein toxin structure. Austin, TX: RG Landes; 1996. p 253–270.
- Monzingo AF, Robertus JD. X-ray analysis of substrate analogs in the ricin A-chain active site. *J Mol Biol* 1992;227:1136–1145.
- Yan X, Hollis T, Svinth M, Day P, Monzingo AF, Milne GWA, Robertus JD. Structure-based identification of a ricin inhibitor. *J Mol Biol* 1997;266:1043–1049.
- Villafranca JE, Robertus JD. Ricin B chain is a product of gene duplication. *J Biol Chem* 1981;256:554–556.
- Rutenber E, Ready M, Robertus JD. Structure and evolution of ricin B chain. *Nature* 1987;326:624–626.
- Girbés T, Citores L, Iglesias R, Ferreras JM, Muñoz R, Rojo MA, Arias FJ, García JR, Méndez E, Calonge M. Ebulin I, a non-toxic novel type 2 ribosome-inactivating protein from *Sambucus ebulus* L. leaves. *J Biol Chem* 1993;268:18195–18199.
- Svinth M, Steighardt J, Hernandez R, Suh JK, Kelly C, Day P, Lord M, Girbes T, Robertus JD. Differences in cytotoxicity of native and engineered RIP can be used to assess their ability to reach the cytoplasm. *Biochem Biophys Res Commun* 1998;249:637–642.
- Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987;162:156–159.
- Wu TH, Chow LP, Lin JY. Sechiumin, a ribosome-inactivating protein from the edible gourd, *Sechium edule* Swartz. Purification, characterization, molecular cloning and expression. *Eur J Biochem* 1998;255:400–408.
- Chenchik A, Diachenko L, Moqadam F, Tarabykin V, Lukyanov S, Siebert PD. Full-length cDNA cloning and determination of mRNA 5' and 3' ends by amplification of adaptor-ligated cDNA. *Biotechniques* 1996;21:526–524.
- Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997;276:307–326.
- Rutenber E, Katzin BJ, Collins EJ, Mlsna D, Ernst SE, Ready MP, Robertus JD. The crystallographic refinement of ricin at 2.5 Å resolution. *Proteins* 1991;10:240–250.
- Brünger AT. Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. *Acta Crystallogr A* 1990;46:46–57.
- Collaborative Computational Project. Number 4: the CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 1994;50:760–763.
- Lunin VY, Skovoroda TD. R-free likelihood-based estimates of errors for phases calculated from atomic models. *Acta Crystallogr A* 1995;51:880–886.
- Urzhumtsev AG, Skovoroda TD, Lunin VY. A procedure compatible with X-PLOR for the calculation of electron-density maps weighted using an R-free-likelihood approach. *J Appl Crystallogr* 1996;29:741–744.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 1991;47:110–119.
- Brünger AT, Kuriyan J, Karplus M. Crystallographic R factor refinement by molecular dynamics. *Science* 1987;235:458–460.
- Brünger AT, Krukowski A, Erickson J. Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr A* 1990;46:585–593.
- Brünger AT. The free R Value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 1992;355:472–474.
- Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997;53:240–255.
- Kleywegt GJ, Brünger AT. Checking your imagination: applications of the free R value. *Structure* 1996;4:897–904.
- Read RJ. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr A* 1986;42:140–149.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26:283–291.
- Monzingo AF, Collins EJ, Ernst SR, Irvin JD, Robertus JD. The 2.5 Å structure of pokeweed antiviral protein. *J Mol Biol* 1993;233:705–715.
- Citores L, Munoz R, De Benito FM, Iglesias R, Ferreras JM, Girbés T. Differential sensitivity of HeLa cells to the type 2 ribosome-inactivating proteins ebulin 1, nigrin b and nigrin f as compared with ricin. *Cell Mol Biol* 1996;42:473–476.
- Van Damme EJM, Barre A, Rougé P, Van Leuven F, Peumans WJ. Characterization and molecular cloning of *Sambucus nigra* agglutinin V (nigrin b), a GalNAc-specific type-2 ribosome-inactivating protein from the bark of elderberry (*Sambucus nigra*). *Eur J Biochem* 1996;237:505–513.
- Rojo MA, Yato M, Ishii-Minami N, Ninami E, Kaku H, Citores L, Girbés T, Shibuya N. Isolation, cDNA cloning, biological properties, and carbohydrate binding specificity of sieboldin-b, a type II ribosome-inactivating protein from the bark of Japanese elderberry (*Sambucus sieboldiana*). *Arch Biochem Biophys* 1997;340:185–194.
- Hatakeyama T, Yamasaki N, Funatsu G. Identification of the tryptophan residue located at the low-affinity saccharide binding site of ricin D. *J Biochem* 1986;100:781–788.
- Frankel AE, Burbage C, Fu T, Tagge E, Chandler J, Willingham MC. Ricin toxin contains at least three galactose-binding sites located in B chain subdomains 1 alpha, 1 beta, and 2 gamma. *Biochemistry* 1996;35:14749–14756.
- Steeves RM, Denton ME, Barnard F, Henry A, Lambert JM. Identification of three oligosaccharide binding sites in ricin. *Biochemistry* 1999;38:11677–11685.
- Araki T, Funatsu G. The complete amino acid sequence of the B-chain of ricin E isolated from small-grain castor bean seeds. *Biochim Biophys Acta* 1987;911:191–200.
- Hatakeyama T, Ohba H, Yamasaki N, Funatsu G. Binding of *Saccharides* to Ricin E isolated from small castor beans. *J Biochem* 1989;105:444–448.
- Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J Appl Crystallogr* 1991;2:946–950.