The three-dimensional structure of ricin at 2.8 Å

(Received for publication, December 15, 1986)

William Montfort†, Jesus E. Villafranca§, Arthur F. Monzino§, Stephen R. Erast§, Betsy Katzin§, Earle Rutenber†, Nuyhen H. Xuong§, Ron Hamlin§, and Jon D. Robertus**

From the †Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, the §Agouron Institute, La Jolla, California 92037, the ‡Clayton Foundation Biochemical Institute, Department of Chemistry, University of Texas, Austin, Texas 78712, and the ††Department of Physics, University of California-San Diego, La Jolla, California 92037.

The x-ray crystallographic structure of the heterodimeric plant toxin ricin has been determined at 2.8-Å resolution. The A chain enzyme is a globular protein with extensive secondary structure and a reasonably prominent cleft assumed to be the active site. The B chain lectin folds into two topologically similar domains, each binding lactose in a shallow cleft. In each site a glutamine residue forms a hydrogen bond to the OH-4 of galactose, accounting for the epimeric specificity of binding. The interface between the A and B chains shows some hydrophobic contacts in which proline and phenylalanine side chains play a prominent role.

A variety of plants contains cytotoxic proteins which are among the most poisonous compounds known. Perhaps the best known of these toxins is ricin, isolated from the seeds of the castor plant Ricinus communis. The protein was first isolated by Stillmark in the 1880s (1) and has been extensively studied since then. A comprehensive review of the voluminous work on ricin was published in 1982 by Olsnes and Pihl (2).

Ricin is isolated as a heterodimer, consisting of a 32,000-dalton A chain glycoprotein linked by a disulfide bond to a 32,000-dalton B chain glycoprotein (3). The B chain is a lectin with a binding preference for galactosides (4), although it binds much more strongly to complex galactosides from cell surface carbohydrates than to simple sugars (5). Binding of the B chain to cell surface receptors appears to trigger endocytotic uptake of the protein (6). Recent studies have suggested that A and B chains have a strong affinity for one another, mediated by hydrophobic forces, and that association of the chains is necessary for toxicity. The disulfide bond between the chains does not appear to be critical for toxicity except in maintaining protein-protein interactions at very low toxin concentration (7). There is recent evidence to suggest that the B chain also facilitates escape of the A chain from endosomes into the cytoplasm (8). These same workers also showed that, in addition to the galactoside binding uptake mechanism, ricin can be taken up by and intoxicate those cells which possess mannose receptors. In this case, the mannose residues of ricin are themselves recognized by a cellular receptor (8).

Once the A chain of ricin reaches the cytoplasm of a typical eucaryotic cell, it enzymatically attacks the 60 S ribosomal subunit and disrupts protein synthesis (2). Ricin has been shown (9) to have a \( K_m = 0.1 \mu M \) for ribosomes and shows a \( k_{cat} = 1500 \text{ min}^{-1} \). Bacterial ribosomes are not susceptible to ricin intoxication (10). The exact nature of the catalytic event is unknown. No cofactors are required, and the reaction is not similar to that of diphtheria toxin which attacks soluble elongation factors. It has been shown that ricin has a very weak ribonuclease activity against rRNA isolated from the 60 S subunit of reticulocyte ribosomes (11). Despite many efforts, there is no evidence as yet for any gross alteration, including rRNA cleavage, of ricin-intoxicated ribosomes (2). It seems reasonable, however, that the ribonuclease activity may be an aberration of normal ricin action and that ricin action may involve the binding of rRNA in some fashion.

In addition to its biochemically interesting properties, ricin has attracted considerable interest as a potential antitumor agent. The protein has been shown to be roughly 10–100 times more toxic to certain transformed cell lines than to normal cells (6). However, the most exciting use of ricin has been in the form of immunotoxins. These are conjugates of ricin with cell-specific antibodies which act to target the toxin (12–14). Although the use of immunotoxins as anticancer agents has been reasonably encouraging, there have been problems due to lack of target cell specificity. It has been observed that if the B chain cell surface binding function is reversibly blocked, as by galactose, then immunotoxin specificity increases (13). This has led to the hope that genetic engineering of ricin could produce more specific therapeutic reagents.

The amino acid sequences and carbohydrate compositions of both ricin chains have been determined by chemical methods (15, 16). More recently, Lord and Butterworth (17) have shown that both chains of ricin are coded by a single mRNA. The nucleotide sequence for cDNA derived from this message has been determined (18). That sequence revealed that ricin is synthesized as a proenzyme, including a leader sequence upstream of the A chain and 12-amino acid peptide linking the A to the B chain. The leader peptide and the linking peptide are both processed away to form the heterodimer. The nucleotide sequence of the genomic DNA has also been determined (19). It confirms the proenzyme structure of ricin and also reveals that the gene has no introns.

In this paper we report the three-dimensional structure of ricin, as determined by x-ray crystallography at 2.8-Å resolution.

** Experimental procedures

Ricin was prepared and crystals grown as described previously (7, 20). The crystals are orthorhombic space group \( P2_12_12 \), and with \( a \)
Phase refinement statistics as a function of resolution. PtTP refers to the 2-hydroxyethane thiolate (2,2',2''-terpyridine) platinum dinitrate derivative. MeHg refers to methyl mercury. $f_i$ is the root mean squared heavy atom structure factor, $E$ is the root mean squared lack of closure, and $fom$ is the average figure of merit.

| Resolution | Number of reflections | PtTP $f_i$ | $E$ | Number of reflections | MeHg $f_i$ | $E$ | $fom$ |
|------------|-----------------------|------------|-----|-----------------------|------------|-----|-------|
| 8.84       | 174                   | 39         | 28  | 164                   | 46         | 32  | 0.88  |
| 6.25       | 1012                  | 35         | 20  | 968                   | 43         | 28  | 0.86  |
| 5.10       | 1285                  | 32         | 17  | 1125                  | 40         | 24  | 0.82  |
| 4.42       | 1502                  | 29         | 17  | 1394                  | 37         | 26  | 0.79  |
| 3.95       | 1670                  | 26         | 16  | 1511                  | 35         | 26  | 0.75  |
| 3.61       | 1842                  | 24         | 17  | 1664                  | 33         | 24  | 0.71  |
| 3.34       | 1990                  | 22         | 14  | 1791                  | 31         | 20  | 0.68  |
| 3.13       | 2111                  | 19         | 13  | 1896                  | 29         | 18  | 0.66  |
| 2.95       | 2234                  | 18         | 11  | 2024                  | 27         | 18  | 0.65  |
| 2.80       | 2242                  | 16         | 10  | 1999                  | 25         |     | 0.62  |

**RESULTS AND DISCUSSION**

Fig. 1 shows a stereo of the $\alpha$ carbon tracing of ricin. The B chain occupies the lower left portion of the figure. It is roughly 70 Å long, about 30 Å wide, and is composed of two separate folding domains. Each domain is a roughly 30-Å sphere and binds one lactose disaccharide, as predicted earlier (27).

The A chain is a more globular protein, situated in the upper right of Fig. 1. The protein has the shape of an oblong disc, approximately 55 Å long, 45 Å wide, and only about 35 Å thick. Fig. 2 is a schematic ribbon drawing of the protein; the A and B chains have been separated slightly for clarity. The carboxyl-terminal 57 or so residues of the A chain fold into a protruding domain that interacts with the B chain by sliding into the cleft between the two B chain domains. A disulfide bond is formed between residues 259 of the A chain and 4 of the B chain. The amino-terminal 8–10 residues of the B chain form a long arm which interacts with the carboxyl-terminal domain of the A chain. As a result, the interchain disulfide is in an exposed region quite far removed from the main body of the B chain.

We now consider each chain in slightly greater detail. The A chain of ricin exhibits a substantial amount of secondary structure. It contains seven $\alpha$ helices totaling about 80 amino acid residues, meaning that about 30% of the protein is helical. There also exists a reasonably pronounced, five-stranded $\beta$ sheet composed of residues 57–116, although not all the residues in that stretch are directly involved in the sheet. In all, the A chain is about 15% $\beta$ structure.

The A chain folds into three somewhat arbitrary domains, as shown schematically in Fig. 3. The amino-terminal 117 residues comprise the first domain (shown in Fig. 3a), which is dominated by the five-stranded $\beta$ sheet. This flat domain forms the bottom of the A chain as we usually view it down the crystallographic $a$ axis. The second domain (Fig. 3b) is composed of residues 118–210 and is dominated by five $\alpha$ helices. This domain generally sits over the first domain and slightly to the left. The only free sulfhydryl group in the protein, Cys-171, lies on the second helix in this domain. As expected, it bonds to the single methyl mercury of that isomorphous derivative. The third domain, residues 211–267 (shown in Fig. 3c), forms a compact, disc-like domain, roughly 25–30 Å across and 15–20 Å thick. This domain interacts with the first two domains and is also involved, as mentioned earlier, in strong interaction with the B chain. Domain 3 appears to be distinct from domain 2 in that residues 186–219 are all helical, but the helix makes a sharp bend at position 211.

Very few details are presently known about the mechanism.
FIG. 1. Stereogram of the ricin Ca backbone. The A chain is in the upper right of the figure and has open bonds. The B chain, at the lower left, is an elongated, two-domain structure. The two lactose molecules bound to the protein are indicated at either end of the B chain. The view is directly down the crystallographic a axis, which minimizes overlap of the two chains.

FIG. 2. Ribbon representation of the ricin backbone. The A chain is in the upper right and the B chain at lower left. The two lactose moieties bound to the B chain are each represented as pairs of discs. The chains have been separated slightly to facilitate viewing. The disulfide bond linking the chains is indicated in the lower right portion of the molecule. The A chain has been divided into three domains where residues 1–117 have a small dot pattern, residues 118–210 are an open ribbon, and residues 211–267 have a large dot pattern.

of A chain action. In fact, even the substrate is unknown, although the large subunit of eucaryotic ribosomes is the target (2). The protein does have a reasonably prominent cleft, however, as seen in the Ca drawing of Fig. 4. It is very likely that this is the active site cleft, or latent active site, since a conformational rearrangement may take place when the A chain is released from the B chain. As yet, no active site residues have been identified by chemical studies. Further work, perhaps by site-directed mutagenesis, will be required to define the role of residues in this vicinity. The putative active site cleft is created at the interface between all three domains. This cleft runs roughly horizontally and downward across the center of the A chain as viewed in Fig. 1. The cleft has good access to solvent in the heterodimer, suggesting that a conformational change may occur when A chain is released from the B chain since the heterodimer itself is inactive.

The B chain is a lectin with an affinity for galactosides. The protein has been known for some time to bind two galactosides in a noncooperative fashion (28, 29). In one study the two sites had different galactoside affinities, showing $K_d$ values of 0.03 and 0.4 mM (28), whereas studies with fluorescent galactosides suggested that the two binding strengths are
FIG. 3. Folding domains of the ricin A chain. To facilitate interpretation of the A chain fold, the structure is broken into three domains. a, residues 1–117 are dominated by a five-stranded β sheet structure. b, residues 118–210 are dominated by five α helices. This compact domain sits directly over the first domain. c, residues 211–267 fold into a compact domain. It interacts with the two other A chain domains, but also makes close contacts with the B chain, fitting into the cleft between the two B chain domains. The view in this figure is directly down the crystallographic α axis. The pattern is consistent with that of the A chain shown in Fig. 3.

FIG. 4. The putative active site cleft of A chain. The Co backbone of A chain has been rotated to expose an extensive cleft which is not clear in the usual view down the crystallographic α axis. As yet, there is no chemical evidence to implicate any residues in the catalytic action of the A chain.

more nearly equal (29). Chemical studies have implicated Tyr-248 and an unknown Trp in the two binding sites (30, 31).

We have previously shown that the B chain is a gene duplication product, showing about 32% amino acid identity between the two halves (27). Each domain is built around a pair of disulfide loops, and we expected it to be a two-domain, each binding one lactose. This is indeed true, and the two domains have identical folding topologies as can be seen in Fig. 2. The amino-terminal domain consisting of residues 1–135 can be related to the carboxyl-terminal domain (residues 136–267) in this view by a rotation of roughly 90° counterclockwise around the axis normal to the page. The lactose binding site of each domain is shown as a pair of circles in Fig. 2.

The native crystals contain the disaccharide lactose, and the two lactose binding sites are quite clearly defined in the native electron density map. (In addition, we have produced a 4.5-Å difference density map between parent and lactose free ricin). The native electron density map for the carboxyl-terminal domain, with the bound lactose and several key residues superimposed, is shown in Fig. 5a. It is clear that even in this unrefined model, key residues can be positioned quite accurately.

The electron density shows that in each domain the galactose moiety of the bound lactose lies in a shallow pocket. One side of the pocket is the aromatic ring of a side group, and the other is a slight kink in a locally straight section of backbone. The rings are those of Trp-37 in domain 1 and Tyr-248 in domain 2. In domain 1 the galactose forms hydrogen bonds to Asn-46 and to Lys-40. These bonds involve OH-3 and OH-4 of the sugar and help determine the epimeric specificity of the binding site. The key Asn-46 residue is held in position by a hydrogen bond to Asp-22. A hydrogen bond
also appears to be formed between Gln-35 and OH-6 of the galactose. The sugar binding arrangement for domain 1 is shown in Fig. 5b. In domain 2, Asn-255 is crucial to epimeric specificity, bonding to the galactose OH-4. The position of Asn-255 is stabilized by a hydrogen bond with Asp-234. A water molecule, bound to Asn-255, forms a hydrogen bond to OH-3 and is in a position analogous to Lys-40 of domain 1. There is no analog of Gln-35 in the domain 2 galactose binding site. The key elements of this binding site are shown in Fig. 5c.

It is interesting to note that the two domains of the B chain are the product of a gene duplication, have similar folding topologies, and bind galactose in similar ways; yet, the residues involved in that binding are not directly related by the gene duplication. The residues binding galactose in the amino-terminal domain are found in the first third of the domain peptide, whereas those of the carboxyl-terminal are in the last third of that peptide. The implication of the two apparently nonhomologous binding sites in two clearly homologous domains for the evolution and function of B chain will be addressed and thoroughly developed in a separate paper.

Having described the topology of the ricin structure, we can now briefly address certain other questions. It is generally considered that the A and B chains are held together by nonpolar forces (7, 32). The measurements of Lewis and Youle (7) suggest that the free energy of this association is roughly -10 kcal/mol. The free energy of moving various chemical groups from an aqueous to a nonpolar environment has been cataloged by Kauzmann (33). The movement of a methane, similar to an alanine side chain, is favorable by -2.6 kcal/mol, whereas movement of a benzene ring is favorable by -4.1 kcal/mol. Together, this information suggests that the ricin A-B interface need involve hydrophobic contacts between a relatively few residues. Our initial analysis shows only a few residues interacting, and they are largely the side chains of prolines and phenylalanines. For example, prolines 7 and 9 of B chain contact Pro-229 of A chain. Also, Pro-260 of B chain, together with phenylalanines 218 and 219, contact Pro-250 of A chain.

It is also known that ricin is a glycoprotein (15, 16). There are two glycosylation sites on the B chain and two potential sites on the A chain, although A chain is often incompletely glycosylated (34). The two carbohydrate moieties at positions 95 and 135 of B chain are very clear in our present map, and density for several sugars can be seen. There is electron density adjacent to each A chain site (10 and 236), but it is very weak, and we hope that improved phases from crystallographic refinement will allow us to visualize those sugars better.

Refinement will also facilitate detailed comparison of the ricin A chain with trichosanthin. Trichosanthin is a single-chain plant toxin, which has an amino acid sequence ~34% identical with A chain (35), for which the peptide backbone has been traced at 4 Å resolution (36). Our preliminary inspection shows that the overall folding pattern is similar to that of A chain, although some features of secondary structure appear to differ.

Finally, it is well known that the A chain of ricin is not enzymatically active, unless it is released from the B chain (2); it is also known that the release exposes a nonpolar area on the A chain which may aid in membrane transport (37). It is quite likely that release from the B chain causes a confor-
mational change in A chain which activates the enzyme function and also alters the nonpolar surface to facilitate membrane binding. We have recently crystallized A chain expressed from a gene cloned into *Escherichia coli* (38). The crystals diffract well, and a comparison between that structure and the A chain seen here in ricin may shed light on this putative conformational change.

Acknowledgments—We wish to thank Vivian Benningfeld and Terri Kolvoord for their help in preparation of this manuscript and Fred Hoffman for his assistance with figures. The platinum compound used in this study was a gift from Dr. H. T. Wright.

REFERENCES

1. Stillmark, H. (1889) *Arb. Pharmacol. Inst. Dorpat.* 3, 59–151
2. Olsnes, A. and Pihl, A. (1982) in *The Molecular Actions of Toxins and Viruses* (Cohen, F., and Van Heyningen, S., eds) pp. 52–105, Elsevier Biomedical Press, New York
3. Olsnes, S., and Pihl, A. (1973) *Biochemistry* 12, 3121–3126
4. Nicolson, G. L., and Blaustein, J. (1972) *Biochim. Biophys. Acta* 206, 543–547
5. Baenziger, J. U., and Fiete, D. (1979) *J. Biol. Chem.* 254, 9795–9799
6. Nicolson, G. L., Lacorbierre, M., and Hunter, T. R. (1975) *Cancer Res.* 35, 144–155
7. Lewis, M. S., and Youle, R. J. (1986) *J. Biol. Chem.* 261, 11571–11577
8. Simmons, B. M., Stahl, P. D., and Russell, J. H. (1986) *J. Biol. Chem.* 261, 7912–7920
9. Olsnes, S., Fernandez-Puentes, C., Carrasco, L., and Vasquez, D. (1975) *Eur. J. Biochem.* 60, 291–298
10. Olsnes, S., Heiberg, R., and Pihl, A. (1973) *J. Biol. Rep.* 1, 15–20
11. Obrijk, T. G., Moran, T. P., and Colinas, R. J. (1985) *Biochem. Biophys. Res. Commun.* 130, 879–884
12. Vitetta, E. S., Krolick, K. A., Munno, M. I., Cusley, W., and Uhr, J. W. (1985) *Science* 219, 644–649
13. Thorpe, P. E., Mason, D. W., Brown, A. N. F., Simmonds, S. J., Ross, W. C. J., Cumber, A. J., and Forrester, J. A. (1982) *Nature* 297, 594–596
14. Vailera, D. A., Ash, R., Zanjani, E. P., Kersey, J. H., LeBien, T. W., Beverley, P. C. L., Neville, D. M., Jr., and Youle, R. J. (1985) *Science* 222, 512–514
15. Funatsu, G., Yoshitake, S., and Funatsu, M. (1978) *Agric. Biol. Chem.* 42, 501–503
16. Funatsu, G., Kimura, M., and Funatsu, M. (1979) *Agric. Biol. Chem.* 43, 2221–2224
17. Butterworth, A. G., and Lord, J. M. (1983) *Eur. J. Biochem.* 137, 50–65
18. Lamb, F. I., Roberta, L. M., and Lord, J. M. (1985) *Eur. J. Biochem.* 148, 265–270
19. Halling, K. C., Halling, A. C., Murray, E. E., Ledin, B. F., Houston, L. L., and Weaver, B. F. (1985) *Nucleic Acids Res.* 13, 8019–8033
20. Villafranca, J. E., and Robertus, J. D. (1977) *J. Mol. Biol.* 116, 331–335
21. Rice, D. W. (1981) *Acta Crystallogr. Sect. B Struct. Crystallogr.* *Cryst. Chem.* A37, 491–500
22. Remington, S., Wiegand, G., and Huber, R. (1982) *J. Mol. Biol.* 158, 111–152
23. Bhat, T. N., and Blow, D. M. (1982) *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* A38, 21–28
24. Hamlin, R. (1985) *Methods Enzymol.* 114, 416–452
25. Wang, B.-C. (1985) *Methods Enzymol.* 119, 90–112
26. Jones, T. A. (1978) *J. Appl. Crystallogr.* 11, 268–272
27. Villafranca, J. E., and Robertus, J. D. (1981) *J. Biol. Chem.* 256, 554–556
28. Zentz, C., Freney, J. P., and Bourrillon, R. (1978) *Biochim. Biophys. Acta* 538, 18–26
29. Houston, L. L., and Dooley, T. P. (1982) *J. Biol. Chem.* 257, 4147–4151
30. Mise, T., Shimoda, T., and Funatsu, G. (1986) *Agric. Biol. Chem.* 50, 555–561
31. Hatakeyama, T., Yamasaki, N., and Funatsu, G. (1986) *J. Biochem. (Tokyo)* 99, 1049–1056
32. Houston, L. L. (1980) *Biochem. Biophys. Res. Commun.* 92, 319–326
33. Kuzmann, W. (1959) *Adv. Protein Chem.* 14, 1–63
34. Foxwell, B. M. J., Donovan, T. A., Thorpe, P. E., and Wilson, G. (1985) *Biochim. Biophys. Acta* 840, 193–203
35. Xuejun, Z., and Jiahuai, F. (1985) *Acta Crystallogr. Sect. B Struct. Crystallogr.* *Cryst. Chem.* A37, 491–500
36. Pan, K., Zhang, Y., Lin, Y., Wu, S., Zhang, M., Chen, S., Xia, Z., Tian, G., Ni, C., Fan, Z., Ma, Y., and Sun, X. (1986) *Science* 230, 26–32
37. Houston, L. L. (1982) *J. Biol. Chem.* 257, 1532–1539
38. Robertus, J. D., Piatak, M., Ferris, R., and Houston, L. L. *J. Biol. Chem.* 262, 19–20